DOCUMENT RESUME

ED 272 582 TM 860 528

AUTHOR	Brandenburg, Dale C.; Smith, Martin E.
TITLE	Evaluation of Corporate Training Programs.
INSTITUTION	ERIC Clearinghouse on Tests, Measurement, and Evaluation, Princeton, N.J.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO	ERIC-TME-R-91
PUB DATE	May 86
CONTRACT	400-83-0015
NOTE	115p.
AVAILABLE FROM	ERIC Clearinghouse on Tests, Measurement, and Evaluation, Educational Testing Service, Princeton, NJ 08541 ($9.00).
PUB TYPE	Reports - Descriptive (141) -- Information Analyses - ERIC Information Analysis Products (071)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	*Corporate Education; Curriculum Evaluation; Data Collection; Decision Making; Evaluation Criteria; *Evaluation Methods; Evaluation Problems; *Formative Evaluation; Human Resources; Job Training; *Labor Force Development; Postsecondary Education; *Summative Evaluation; *Training Methods; Training Objectives

ABSTRACT

The growing interest in human resources development (HRD) is a response to the increasingly competitive business environment which, in turn, has been shaped by such forces as the emergence of global competition, the quickening pace of technological innovation and the reduction of federal regulations in some industries. This monograph seeks to contribute to the current momentum of HRD by describing the state of the art for HRD evaluation. The needs of business managers and academicians are addressed by describing the applications of evaluation methods to decision-making within the training or HRD organization. Both formative and summative training evaluations are discussed in terms of issues, constraints, and techniques. Models for training evaluation are analyzed into a number of components: (1) definition of perspective; (2) description of variables; (3) use of flowcharts or descriptive matrices; (4) lists of questions; (5) data collection techniques; (6) data analysis techniques; (7) research designs; (8) reporting procedures; and (9) detailed examples. A section on evaluation practice summarizes analyses of evaluations to show how training evaluations are actually done. Conceptual and strategic issues that challenge the evaluator are described. An appendix outlines methods for training economic impact analysis. (LMO)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED272582

EVALUATION OF CORPORATE TRAINING PROGRAMS

TME REPORT 91

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

by

Dale Brandenburg
University of Illinois

and

Martin E. Smith
New England Telephone Company

ERIC®

TESTS MEASUREMENT EVALUATION

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY 08541-0001

TM 840 528

ERIC/TME REPORT 91

EVALUATION OF CORPORATE TRAINING PROGRAMS

DALE C. BRANDENBURG

University of Illinois

and

MARTIN E. SMITH

New England Telephone Company

MAY 1986

ERIC Clearinghouse on Tests, Measurement, and Evaluation.
Educational Testing Service, Princeton, New Jersey 08541-001

The material in this publication was prepared pursuant to a contract with the Office of Educational Research and Improvement, U.S. Department of Education. Contractors undertaking such projects under government sponsorship are encouraged to express freely their judgment in professional and technical matters. Prior to publication, the manuscript was submitted to qualified professionals for critical review and determination of professional competence. This publication has met such standards. Points of view or opinions, however, do not necessarily represent the official view or opinions of either these reviewers or the Office of Educational Research and Improvement.

ERIC Clearinghouse on Tests, Measurement, and Evaluation
Educational Testing Service
Princeton, NJ 08541

This publication was prepared with funding from the Office of Educational Research and Improvement, U.S. Department of Education under contract No. NIE-400-83-0015. The opinions expressed in this report do not necessarily reflect the positions or policies of OERI or the Department of Education.

TABLE OF CONTENTS

	Page
INTRODUCTION	1
I. Evaluation Defined	2
II. Intended Readers	4
III. Content Overview	5
TRAINING EVALUATION APPLICATIONS	7
I. Formative Trainee Evaluation	10
II. Summative Trainee Evaluation	11
III. Formative Course Evaluation	13
IV. Summative Course Evaluation	16
V. Formative Curriculum Evaluation	20
VI. Summative Curriculum Evaluation	22
VII. Formative Training Organization Evaluation	25
VIII. Summative Training Organization Evaluation	28
IX. Formative Corporate Evaluation	30
X. Summative Corporate Evaluation	32
GENERAL EVALUATION MODELS	37
I. Model Characteristics	39
II. Summary of Models	41
PRACTICE	57
I. References	57
II. Findings	58
ISSUES	67
I. Client's Commitment	67
II. Evaluating the Training Organization.	70
III. Cost Benefit Analyses	76
IV. Evaluating Management Training	85
V. Evaluating Commercially Available Training	90
REFERENCES	97
APPENDIX	105

INTRODUCTION

Nineteen eighty-six is an exciting time for trainers. In the past, training has generally been treated as a corporate overhead, susceptible to budget cuts, often a "dumping ground" for out-of-favor managers, not worth the attention of senior management. Trainers reinforced their status by uncritically adopting fads unrelated to corporate needs.

Times have changed. The popular business press is dominated by books on human resources, for example, In Search of Excellence, The Change Masters, The One Minute Manager. Many companies see human resources as the critical asset that enables a competitive edge. The private sector is investing huge sums in human resources development. In 1981, GM, IBM and AT&T reported training expenditures of $1 billion, $750 million and $750 million, respectively (Thomas, 1981).

The growing interest in human resources development (HRD) is a response to the increasingly competitive business environment which, in turn, has been shaped by such forces as the emergence of global competition, the quickening pace of technological innovation and the reduction of federal regulation in some industries.

Societal trends have also focused attention on human resources issues. Younger workers, especially the college educated, tend to find a life style first and then a job to support it (Odiorne, 1980). Supervisors tend to value work more highly than their younger employees. The difference in values creates the potential for conflict. Other work force trends include growing proportions of women and minorities in the workforce and in management positions, an increasing number of bilingual workers with English as their second language, an increasing number of

two-income families who are less likely to accept geographical transfers, legislation protecting the rights and health of employees, and so on.

Amid these business and societal trends, HRD is evolving from an "arts and crafts" mode toward a technology. We see systematic approaches for analyzing work and organizational problems, instructional design, job and organizational design, the design of incentive and feedback systems. In many companies, training is part of an integrated array of human resources services. Trainers are also developing smarter political skills for getting HRD into the corporate mainstream.

There is an energy about HRD which we see influencing academic curricula, research, publications and professional societies. This monograph seeks to contribute to that momentum by describing the "state-of-the-art" for a function critical to HRD: evaluation.

I. Evaluation Defined

We offer our definition of the term "evaluation". Evaluation is a judgment of an entity on some dimension valued by the client. This judgment is based upon a measurement of actual status on the dimension against a standard. The underlined terms deserve further explanation.

The term "judgment" denotes an appraisal or a decision. Judgments are generally divided into two classes: summative and formative. Summative evaluations are directed at "go/no go" decisions, for example: Should we purchase the XYZ program? Is the new course better than the one it replaced? Audiences for summative evaluations are often the senior management of the client organization(s) because they will decide whether or not to send people to the course. In contrast, formative evaluations seek to identify ways of improving the evaluated entity; for example: If

XYZ is purchased, what modifications will be needed? What parts of the job are XYZ graduates not able to do satisfactorily? Trainers are generally the audience for formative evaluations because they are the people who will fix the course. The same study may support both summative and formative purposes.

"Entity" refers to the object of an evaluation. Generally, we think of evaluating a training course, but any aspect of the training or HRD function is fair game for an evaluation, including: trainees, instructors, instructional strategies, facilities, the training organization itself. Organizational interventions, with little or no instructional component, can be evaluated, including: feedback and incentive systems, team-building, MBO processes, personnel selection and placement, organization and job design projects.

The evaluation "dimension" is a critical characteristic of the entity that is valued by the audience for the evaluation. For example, an evaluation of a training program might seek to determine the relevance of program content to the job requirements of the target population, or the efficiency of the development process, or the timeliness of delivery process. Program evaluations most often concentrate on the outcomes of training, which may be grouped into four categories, (Kirkpatrick, 1976): participant reactions, learning, job proficiency, and the organizational consequences of the graduates' job proficiency.

"Measurement of actual status" is defined as collecting data to show how "things really are" (or, at least, how people think they are), in terms of the evaluation dimension. Measurement techniques may be divided into several categories: opinion data gathered by interview or questionnaire, knowledge testing, performance testing under simulated job

conditions, observation of job performance, and organizational measures of job performance, global measures of the organization's performance. Specific measurement techniques will be discussed later.

A "standard" is a criterion for judging success or failure. Standards may be relative (or norm-referenced), on one hand, or absolute (or criterion-referenced), on the other hand. If we use the context of program evaluation, then a relative standard implies a second sample of data collected from: the trainees prior to the program (i.e., pretesting); employees who receive an alternative, perhaps a program that is being replaced by the program being evaluated; or a combination of these conditions. Absolute standards are set independently of any evaluated group's performance. Absolute standards can be derived from: corporate or departmental policy; historical records for measured jobs; conventional practice (e.g., the "90/90" criterion once popular for developing programmed instruction texts); a panel of job experts; or the decision of the evaluation. Needless to say, the same study may employ both relative and absolute standards.

II. Intended Readers

This monograph is directed at two audiences. The first consists of managers in business and industry who have no background in evaluation but who are familiar with how job performance is measured and with organizational factors that may influence the evaluation of training. We assume that these managers are reading this monograph for guidance on how to (a) design an evaluation study, (b) oversee a consultant hired to evaluate a training program, or (c) establish evaluation as a regularly occurring function of the training organization.

The second consists of academicians who are experienced in evaluation methodology but who are unfamiliar with the problems and opportunities of industrial applications of evaluation methodology. We assume that academicians want information on (a) potential uses of their skills in business and industry, (b) problems commonly encountered in training evaluation, (c) research issues, and (d) potential biases of business clients who might support evaluation activities.

III. Content Overview

We will attempt to meet the needs of these two audiences by describing (a) the applications of evaluation methods to decision-making within the training or HRD organization, (b) models proposed for training evaluation, (c) prevailing practice in terms of design features, data collection techniques, and commonly encountered problems, and (d) conceptual and strategic issues that challenge the evaluator.

The reader should recognize that this monograph is not a manual on evaluation design, nor is it a comprehensive literature review. The writers also made only minor attempts to compare the status of training evaluation to that of educational evaluation.

TRAINING EVALUATION APPLICATIONS

A continuing reader of professional magazines such as Training or the Training and Development Journal is often confronted with numerous articles on training evaluation. If a personal reference file was made of such articles, it would be difficult to find a classification scheme sufficient to be usable. Why? Because each writer is thinking of or referencing a particular application for evaluation, and the reader is left to supply the "big picture". What is the big picture? That's the purpose of this section. With the addition of examples spread throughout to facilitate understanding, our purpose is to build a classification scheme worthy of your files as well as to enhance your repertoire in analyzing evaluation applications.

Consider performing training evaluation to answer the following questions:

1. Will training materials be effective when implemented?
2. Does the return on training justify its continuation?

In general, data to answer the first question is easier to get than the second, but that is not the point of this illustration. Using the terms defined earlier, the first is a formative question and the second is a summative one. These two types of judgments give us an initial means to examine a classification of evaluation. A second means is the entity or object of judgment. The first question concerns training materials while the second concerns the entire training organization. This two-way classification, using type of judgment and entity, becomes our "road map" for understanding the intended purposes of evaluation.

The chart of Table 1 (page 9) displays our classification scheme. By Formative Evaluation, we mean activities that are conducted primarily to support or control functions within the training organization itself. Such activities are oriented toward feedback and developmental assistance used mainly by training developers and their managers. Most of these activities occur prior to full field implementation of a training project.

By Summative Evaluation, we mean activities that are conducted after field implementation to support or confirm training goals established during the formative stage. Results of such evaluations often are not only used by training managers, but also by line and senior management external to the training organization. Activities are oriented toward the "impact" of training--productivity improvement, cost savings, return on investment, and the like.

The second dimension of our two-way classification is to simply examine "what is evaluated" (the entity): Trainee, Course, Curriculum, Training Organization, Corporation. The first two of these are self-explanatory. By Curriculum, we mean a series of interrelated courses or other planned learning experiences designed around a single topic for a single target population. For the Training Organization, we are concerned with all products and services emanating from the training unit. The fifth category includes those aspects of the organization that are primarily strategic or affecting the entire organization. Specific training activities may be part of this activity, but it is generally directed at larger organization interventions.

TABLE 1: CLASSIFICATION OF EVALUATION APPLICATIONS

What purpose does evaluation serve?

What Is Evaluated?	FORMATIVE		SUMMATIVE	
	Topic Labels	Example Questions	Topic Labels	Example Questions
TRAINEE	I Diagnosis Remediation Skill Building	• Are course participants achieving the objectives? • What are participant strengths and weaknesses?	II Certification Trainee Proficiency Mastery Job Performance	• Can demonstrated learning gains be observed or measured on the job? • What level of performance distinguishes experienced workers from newly trained workers?
COURSE	III Development Instructor Evaluation Vendor Course Evaluation Unit Evaluation Immediate Follow-Up End-of-Course Performance	• Will the materials be effective when the development process is completed? • Will course meet identified needs? • Is the instructor effective in teaching this course? • What are the gains in learning?	IV Course Proficiency Validation Field Testing Cost Effectiveness Cost Benefit	• Does return on investment justify continuation of this course? • What are the true costs and benefits of this course? • Has this course eliminated a training need?
CURRICULUM	V Program Development Program Improvement Monitor, Record Keeping on Courses	• Is the program consistent with organizational targets for training interventions? • Are the courses sufficiently integrated? • What are the costs of training?	VI Field Effectiveness Impact Evaluation Cost Effectiveness Cost Benefit	• Is this program effecting operating units P & L thru yield enhancement, cost savings, etc.? • What training outcomes relate directly or indirectly to improved job performance?
TRAINING ORGANIZATION OR SYSTEM	VII Client Satisfaction Training Performance Audit (Management Audit)	• Are line management needs likely to be met? • Can the proposed training be integrated with other current corporate activities?	VIII Performance Appraisal System Impact Program Audit	• Are solutions to business problems implemented and reinforced for the projected life of programs? • Is the training system providing an acceptable rate of return on investment?
CORPORATE	IX Front-End Analysis Consistency of Employee Skills with Organizational Mission Management Audit	• What needs can be met by the implementation of a training intervention? • What job behaviors can improve attainment of organizational targets?	X Culture Change Organizational Performance Audit Contribution to Organizational Profit and Loss	• Are mechanisms in place to provide sufficient feedback for projecting future human resource requirements? • Are prescribed organizational targets consistent with the mission of the organization?

I. Formative Trainee Evaluation

Description

The formative evaluation of trainees may also be labelled as "control of student progress". The major objective of activities conducted under this category is to determine if trainees have satisfied or achieved the learning objectives of a training activity--this is the data collection implied by Kirkpatrick's (1976) "Learning" stage. Evaluation of trainees takes place prior to course attendance as well as during attendance.

One evaluation question posed by Alden (1983) relates to prerequisites: "Are the prescribed entry requirements for the program associated with success on the training outcomes?" In other words, do prerequisites match achieved learning? Thus, proper placement of trainees into a sequence of instruction is worth evaluating to maximize use of a trainee's time as well as increase the efficiency of the training process.

Once the trainee enters a program, we are interested in measuring ongoing progress to (a) choose or modify subsequent learning experience and (b) prescribe remediation of individual deficiencies. Consequently, such measures can provide feedback to the trainee for directing advanced or remedial study. A flowchart depicting these activities is provided by Lyons (1972).

Formative evaluation generates information to aid in choosing among various options. Diagnosis of trainee strengths and weaknesses is contingent on the precision of the diagnostic instrument. Such precision is necessary to guarantee that the process of initial placement, selectively correcting a trainee, and advancing the trainee to the next segment only when learning objectives are achieved, can be feasibly

implemented. This type of evaluation does not necessarily take into account how well learned objectives apply on the job--this is part of the trainee summative evaluation.

Techniques

Typical data collection techniques for formative trainee evaluation include the following:

- paper and pencil in-class achievement tests (multiple-choice, essay, short answer)
- commercial placement tests
- simulation exercises
- video tapes, interviews
- checklists

The decisions made following the results of such measures, e.g., whether to recycle a trainee back through a certain sequence, should be weighed against the risks of making a wrong decision. Precision, including validity and reliability, is important when we might be wasting an employee's time--however, sending an incompetent employee to the next learning sequence may have even more important risks. Considering the consequences of such actions, it is worthwhile to note that this topic of evaluation receives little emphasis in the training literature.

II. Summative Trainee Evaluation

Description

In the summative evaluation of trainees, we are primarily concerned with on-the-job performance once a trainee leaves the training environment. It is not so much a concern about whether or not a trainee achieved the prescribed learning objectives, but whether these were

transferred into measurable performance improvement. The major questions are those of certification or verification: "Do graduates of the program exhibit mostly performance under normal job conditions after a practical period of on-the-job experience?" (Alden 1983). Amount of training is not important; demonstrated proficiency is.

Techniques

There exist two basic arenas for collecting data related to summative trainee evaluation--job behavior and job results or accomplishments. Most of the following techniques may be used for these two topics:

- o interviews with line management
- o observations by line managers or other personnel
- o performance appraisals
- o experimental research designs
- o follow-up interviews and self-reports with trainees
- o external expert review
- o case study
- o other document reviews (e.g., feedback from external resources, absenteeism)

Issues and Constraints

One of the most difficult aspects of measuring individual performance change is whether or not one can attribute the results directly to a training course or program. The situation is probably clearer in skill-based training vs. management training, but the proof of cause and effect is difficult, if not impossible, to come by. The working environment is subject to change, business conditions change, workers are transferred--these and other situations inhibit the collection of valid data.

A second concern is the timing of measurement of change. Is two weeks long enough? Three months? One year? The best probable answer: it depends. Time from training completion to follow-up should be long enough for training effects to occur yet short enough to provide feedback for training revisions. Thus, there is a dependency on the type of training as well as the need for feedback.

A third issue revolves around line management support for training. Opportunity to apply learned techniques and the encouragement for doing so are management prerogatives. Involving managers and supervisors in the evaluation effort is a good way to obtain credibility with line employees as well as to enhance the quality of data collection (Dopyera & Pitone 1983).

The quality of measuring instruments leads to the issue of EEO or affirmative action guidelines. Tests must be shown to be job-related; performance data must be collected carefully. Legal issues surrounding testing employee strengths and weaknesses virtually suspended such operations in training organizations in recent years.

Another problem is related to access to employees following training. Employee transfer, commitment to training evaluation (Thompson, 1978) are two sub-issues. These conditions virtually doom any attempt at mailed questionnaire follow-up.

III. Formative Course Evaluation

Description

Almost all formative course evaluation, or quality control in Foshay's (1984) terms, is directed to provide information for use by internal training staff for improvement in the design, writing, and tryout of a

given training course. In terms of the typical phases of instructional design, formative evaluation occurs at the "design" and "development" stages prior to field implementation. Because this activity is viewed as support to training staff, it is usually conducted by internal personnel, generally an instructional developer trained in the techniques of evaluation. In planning evaluations of this type, the following components are typical considerations (Cummings 1984):

- specification of the business problem the course is designed to address
- course goals
- progress to date, scheduling of first pilot
- the format of how the course is to be delivered
- the instructional support materials (guides, references)
- trainee audience (prerequisite activities or skills)
- identification of data collection techniques
- design of reporting results

Techniques

It is probably fair to say that the techniques used in collecting formative course evaluation are limited by the creativity of the evaluator and developer. Among the common techniques mentioned most often in the literature, one is likely to find the following:

- content analysis by subject matter experts (SME's)
- participant reaction questionnaires
- in-class achievement tests
- classroom observation
- participant interviews, debriefings
- videotape of classroom, participants
- instructor opinions or interviews

Issues and Constraints

In many organizations, this type of application, formative course evaluation is virtually the exclusive training evaluation activity. There are a number of reasons for this to be the case: (a) it is the only application where almost all components can be controlled within just the training organization; (b) it is probably the easiest way to obtain data; (c) it helps in the making of important and timely decisions; and (d) it can be easily cost-justified. Decisions that are made are in direct support of training staff and management and all data collected can be utilized internally. Thus, the relevance of evaluative information serving this training support function is high, and its use is immediate.

One of the primary constraints facing evaluation at this stage is the limitation of time and resources available to complete a detailed evaluation report. Time constraints are so severe in many circumstances that only a cursory evaluation can be conducted. Only when the course is out in the field do people realize the serious problems that existed with the materials or instructors. Once the course is in field implementation, it is very difficult to pull it back and recycle it for further development. Thus, a very finely tuned system for cooperative work between evaluators and developers as suggested by Cummings (1984) yields high quality results.

A second issue is that formative course evaluation for the most part is probably conducted more by instructional developers than by evaluators, as Smith (1982) pointed out. Skills of such evaluators are likely to need refinement. Poor quality evaluation at this stage is likely to lead to serious problems later in the field. Of course, many of the problems

really cannot be dumped on the instructional developer doing evaluation. It is likely that there are many other demands in staffing priorities that require consideration. An evaluation may not seem all that important.

The third issue involves the amount of data collected by such a process. Evaluators sometimes complain that their results have no impact. This situation may, in fact, be due to just the sheer amount of information collected and its relevance to the developers. The criticality of the feedback needs to be specified as well as report length. Just the volume of data can deluge and stall revisions in a program. These and other points are raised by Markle (1979) in her well-known article, "Evaluating Instructional Programs: How Much is Enough?".

IV. Summative Course Evaluation

Description

In summative evaluation of courses, one attempts to get at the "results" of training in Kirkpatrick's (1978) terms. To many professional evaluators, this is "where the action is" in terms of evaluation. Reports from summative evaluation can lead to course continuation or discontinuation; trainers can get solid information on the "real payoff" of their efforts; impact of a course on the organization can be determined; and cost-benefits can be forwarded to senior management. In short, summative evaluation attempts to ferret out and describe those aspects of a course that make a real difference in field implementation.

Techniques

The strategies and methods used to collect summative course evaluation tend to have a higher degree of internal validity than the techniques used in formative course evaluation. This is not to say that formative techniques are "softer", but that more rigor is generally applied to the techniques when used summatively (Parker, 1984). The more popular techniques and methods include:

- surveys and questionnaires
- experimental studies
- quasi-experimental studies
- structured interviews
- systematic expert judgment
- multiple baseline studies
- checklists
- case studies

Examples

It is worthwhile to review some summative evaluation studies reported in the literature to examine how "impact" has been measured. Table 2 (page 18) examines at a cursory level a number of such studies (most reported by Smith and Corbett, 1977).

Issues and Constraints

The threats to conducting a successful summative evaluation of a course are many. It is not uncommon to find that a summative evaluation is conducted by an external consultant. This situation yields a whole set of conditions which may influence the outcome of an evaluation while satisfying some criterion of objectivity. Consider a development effort

TABLE 2

REFERENCES DESCRIBING PRICING FORMULA

REFERENCE	JOB	PROGRAM	BENEFITS	MEASUREMENT	FORMULA
Allen, Amacher and Yaney (1976)	Machine Operators	Re-structure of work group	Reduced turnover and, hence, reduced cost of replacing skilled employees.	Factory turnover was monitored before and after creation of "unit work teams."	(Turnover before change — turnover after change) X employees X average replacementment cost.
Fauley (1975)	Supervisors	Speed reading course	Improved reading speed which makes time available for other activities.	Pre— and post— course reading speeds are compared.	Supervisors X hours saved per day X days X hourly wage.
Jones and Moxham (1969)	Sewing Machine Operators	Initial job training	Productivity gain due to less time to reach standard reduced turnover, and increased efficiency.	Productivity and turnover of trainees compared to employees hired (a) before program was implemented and (b) after it was halted.	$(P_2R_2-P_1R_1)V-(W_2R_2-W_1R_1)$ Where: P = productivity R = retention time with company V = unit value of output W = wages Subscripts 1 and 2 represent "before" and "after" the program, respectively.
Mikesell, Wilson and Lawther (1975)	Tax Auditors	Tax audit workshop	Identification of tax payers' errors which led to more tax revenue for state.	Trainees were compared to their audit division before and after the workshop.	(Post minus pre-workshop revenue for trainees)—(Post minus pre-workshop revenue for audit division).
MacFarland and Keeler (1975)	Construction Job	Initial job training	Reduced time to reach standard and reduced supervisory time.	Foreman rating of recent graduate vs. crew in terms of productivity and time spent coaching.	Productivity gain = trainees X work hours X hourly wage X .5; time = foremen X hours saved X hourly wage.
Smith (1975)	Foremen	Attendance administration course	Reduced absence rates.	Work groups' absence rates tracked for foremen who were trained vs. foremen who were not trained.	(Predicted absences — actual absences) X daily wage.
Rosentreter (1979)	Production Managers	Management training program	Reduced employee turnover (3 others not significant).	Turnover measured before and after for treatment and control groups.	Turnover expected–Turnover observed X employees X replacement cost.
Cullen, Sawzin, Sisson and Swanson (1976)	Semi-skilled Machine Operators	Structured vs. unstructured (OJT) training	(1) decreased time to achieve job competency (2) decreased scrap (waste) (3) increase in solving problems	(1) mean to achievement in hours (2) scrap in lbs. per worker (3) simulation of problems	(1) none (2) scrap for group S — scrap for group U X employees X cost of material per lb. (3) none
Zigon (1984) (Samples Only)	Truck Terminal Managers	Management training	(1) decreased cost per bill (2) increased bill/hr., lbs./hr., labor cost/hr. (3) reduced overtime	(1) bill count and cost (2) bill, lbs., labor (3) amount of overtime	(1) compare vs. previous (2) compare to previous (3) compare over 5 months
Reber and Wallin (1984)	Skilled & Semi-skilled Manufacturing Employees	Safety training	Increased safety performance.	Questionnaire observation of performance.	Compare over four time periods; overall injuries decreased from 84.8/100 to 55.1/100.

that cost a organization over $1 million and 18 months to get into the field. Should the training project director also be the summative evaluator? What would the corporate training manager do in attempting to justify this expenditure to senior management?

As an alternative procedure, many training organizations (Cummings, 1984; Smith, 1982[c]) have founded evaluation units that serve as semi-external consultants for summative as well as formative evaluation. Vested interest in any project is kept to a minimum, corporate allegiance is assured, credibility with line management can be enhanced, but senior management may sometimes need convincing.

These evaluation service units are likely to be found in larger training organizations, but the constraints involving summative evaluation are mostly invariant to organization size. The following list of issues and constraints paraphrased from Parker (1984) relate to conducting summative evaluation studies.

- familiarity and understanding of the organizational milieu
- evaluator credibility and interpersonal skills
- those from whom data is collected may suspect that they, themselves, are the object of evaluation
- need for a credible liaison for external evaluators
- the sheer amount of time an evaluator is likely to spend
- corroborating evidence from numerous sources
- modifications to training "discovered" at local sites
- lack of interest from local management
- time lag in showing results
- lack of access to appropriate personnel

- an environment that limits application of findings
- lack of knowledge on how to conduct studies

The above list is not exhaustive, but it certainly provides some clues as to why follow-up efforts often fail.

V. Formative Curriculum Evaluation

Description

In this section, we differentiate evaluation of curricula from the evaluation of courses. While similarities between the two are many, their differences pose some unique problems.

A curriculum shall be defined as two or more planned learning experiences having a defined sequence or framework designed for a specific target population. Alternatively, Schumacher, Kulp and Childs (1975) defined a curriculum as "a sequenced description of learning (or training) experiences which, when followed, will enable students to perform some aspect(s) of their responsibilities in a functional area". It is important to note that this latter definition keys in on "functional area" rather than on specific job tasks. Thus, one would expect a training curriculum to cut across a number of job titles pertaining to a particular functional area. For example, in a training curriculum for engineers on statistical quality control, there could be courses on introducing problem solving techniques with statistics, the basic use and calculation of statistics and the planning of statistical experiments. Its target group would likely include new engineers as well as senior managing engineers.

Curriculum development is different from course development in the sense that we are probably not interested in the detailed analysis of a single job; rather we place the emphasis on analysis of a broad functional

area. Defining a curriculum requires more than information on detailed job requirements. Instead, the latter is the domain of specific course development.

The implications of this difference for formative curriculum evaluation include the following:

- o developers should be concerned about the structure for satisfying functional area job requirements
- o users of the curriculum (field managers and training administrators) should be able to sequence course offerings to maximize staffing requirements
- o utility of a curriculum should extend to defining the kind of personnel who can accomplish broad job requirements
- o the curriculum should serve as a model against which present training may be compared

Techniques

In addition to the techniques previously stated under Formative Course Evaluation, the following list of methods is suggested:

- o logical analysis of sequences for learning
- o correspondence of learning sequence to application (use)
- o cluster analysis of hierarchical training objectives
- o mathematical modeling for delivery efficiency
- o classification matrices for training objectives vs. job function requirements
- o external judgments (subject matter expert review) of cohesiveness of goals, objectives and methods of instruction
- o review of criteria list for curriculum outcomes

A fine example of the use of hierarchical cluster analysis to formative curriculum evaluation is given by Kettenring, Rogers, Smith and Warner (1976). Additionally, a model for curriculum development is supplied by Schumacher, Kulp and Childs (1975).

Issues and Constraints

It is probably fair to state that the constraints and issues identified for course evaluation acquires additional concerns where a curriculum is concerned. One of these issues is coordination of efforts. Brenden (1981) describes a system for coordinating instructors teaching basic skills to manufacturing apprentices. His method involves a hierarchy of evaluations from identifying general concerns to classifying particular strengths and weaknesses of individual instructors. The overall goal is to obtain integrated and uniform quality of instructional delivery.

A second example provided by Fiero (1984) describes the attempts to put together four courses for engineering training. Two courses were purchased from an external vendor and two courses were developed internally. The vendor courses had to be initially taught by the vendor and no modifications could be made. The internal courses had to use the basic premises of the vendor courses in order to provide integration of the topics. Coordination of instructors, course topics, course sequence, and site independence were only a few issues she had to deal with.

VI. Summative Curriculum Evaluation

Description

Differences between summative course evaluation and summative Curriculum Evaluation are similar to differences at the formative level. That is, we need to shift gears to one stage of generality above the course level. While Summative Curriculum Evaluation is also concerned with job performance, our attention is at the functional level rather than at the task level.

In reviewing the available literature, it appears fair to state that summative curriculum evaluation is not a popular activity. One might think that "program evaluation" would imply a set of courses, but the term is generally synonymous with course evaluation. Perhaps training organizations do not view a set of courses about a topic as a curriculum, and thus do not evaluate them in that light. Planned curricula do not occur near as frequently as planned courses. Regardless of explanation, summative curriculum evaluation is a rare topic.

Techniques

To illustrate a method for summative curriculum evaluation that concentrates on the functional, rather than task, level is the description of a quality control curriculum implemented by Motorola (Fiero, 1984). This Total Quality Improvement (TQI) curriculum was comprised of four courses (two vendor-purchased and two internally developed) and was targeted for engineers and engineering managers. Five sites were studied during the evaluation that had completed training anywhere from three months to nine months earlier. Five basic general questions guided the evaluation:

- Why did the site participate in TQI?
- How was TQI implemented at the site?
- What material was learned during TQI courses?
- Was TQI used on the job and how?
- What was the impact of TQI at the site?

Note that these questions do not focus on individual task performances; rather the focus is on general contribution to overall individual and group performance. Techniques for data collection involved individual and group interviews, observations at work stations and at meetings, document reviews and demonstrations by participants.

Two other examples of curriculum evaluation are those reported by Moore (1984) and Hand & Slocum (1972). Moore's curriculum was a self-development program for clerical and secretarial employees at Massachusetts Mutual Life Insurance Company. Courses included management problems and pressures, transactional analysis values and behavior, job analysis, interpersonal communication, and others. A competency-based approach was used over 12 competency factors measured by ratings from supervisors and trainees. Significant gains on pre-/post-measurement were found for most factors.

Hand and Slocum's experimental study was conducted to determine effectiveness of a managerial human relation training program for changing attitudes and the effect of these attitude changes on the organization. Experimental and control groups were taught in 90-minute sessions once a week for 28 weeks. Essentially, one course was on managerial styles and the second was experimental learning. Three of five criterion measures showed significant results in favor of the experimental group.

The following list summarizes those techniques used in the preceding evaluations and suggests those of potential value:

- experimental design
- quasi-experimental design
- interviews (individual, group, structured, unstructured)
- performance appraisal data
- case studies
- cost-effectiveness analysis
- cost-benefit analysis
- checklists
- multiple baseline studies

- o document review
- o surveys and questionnaires

Issues and Constraints

Many of the comments made in the preceding section on Formative Curriculum Evaluation also apply here; i.e., time constraints, sequencing and coordination of efforts. It may be useful, however, to point out three general types of variables with which one should be concerned. These are:

- o educational variables
- o operational variables
- o organizational variables

A curriculum evaluation plan utilizing these issues stands a greater chance for obtaining the desired outcome.

Educational variables take into account the actual curriculum materials to be evaluated. Here we are concerned with, for example, whether or not the design of sequenced materials makes a difference in the field settings. Additionally, we need to know how new curriculum materials interact with existing courses. For operational variables, we should pay attention to the actual delivery of curricula at a particular site. The number of people, budget, capabilities of field staff, time constraints, and relationship to unit business plans are a few of these variables. Between them, educational and operational variables control the critical scope and overall approach to a curriculum evaluation effort. Organizational variables, on the other hand, determine how much reliance is placed on the other two sets of variables. For example, field reputation of the training unit is an organizational variable that must be taken into account in the design for evaluation. It also works the other

way around; that is, the reputation of the client groups should also be noted. Other organizational variables include senior management styles, control over training, basic philosophy of the business unit and mechanisms for communicating and solving business problems.

VII. Formative Training Organization Evaluation

Description

In general, the formative evaluation of a training organization or system in general attempts to answer what is. It implies the formation of an information system that allows for the flow of needs from line staff to senior management through the training organization. In looking at the operational elements of the total corporate organization, one could say that the training system receives performance requirements that must be met with trainees and instructors while operating under policies and procedures with a given budget.

Techniques

One model for examining training organizations formatively is provided by Lyons (1972). He implies the existence of six essential elements to derive a quality control system for training:

1. Training objectives or performance requirements
2. Proficiency and diagnostic measures
3. Data reduction and analysis
4. Procedures for decision and corrective action
5. Communication procedures
6. Managerial support

The use of training objectives is a keystone for the system and should represent the mission for the training organization. The complete system is designed to act rather than react to organizational plans.

One way to go about collecting similar information is found in a model proposed by Smith (1982) which describes the formulation and operation of a training information system. Major components of this information system are: (1) data bases; (2) output; (3) input. The data bases have four main components: (a) training staff; (b) courses; (c) facilities; and (d) employees. The outputs of the system include delivery and development costs, administration, expenses, and trainee costs. The input section contains information on instructors, students, and the project status. Other outputs of the system include data for long range forecasting and evaluation of courses individually as well as by groups. Outcomes of this system included finding that training costs were higher than anticipated; that some course designs were faulty; that projections were needed for sharp increases in training volume; and that there was a lack of standard procedures for cost accounting, quality control and forecasting.

A third alternative methodology for formative system evaluation is the management audit, which examines the use of resources and addresses the main criterion of efficiency. Such an audit as described by Rothwell (1984[a]), is meant to compare the use of resources such as capital or time to norms involving organizational policies, industry averages, common business practice, or published research findings. These studies emphasize relationship between organization input and output. They attempt to compare the present conditions--the "what is"--to desirable criteria--the "what should be." Such an audit is designed to investigate

complex issues or decisions. It is not meant to be used for attacking problems that require immediate attention.

VIII. Summative Training Organization Evaluation

Description

A very obvious question to ask at this point is: if we have done summative evaluation of trainees, courses and curricula, does this then add up to a summative evaluation of the training system? The not so obvious answer is no. What is missing in stating the negative answer to that question is that trainee, course, and curricula evaluation are only components of the system. They do not necessarily imply all of the products and services that are the function of the training organization or department.

The objective of training system summative evaluation is similar to a form of auditing in attempting to isolate effects of a training organization from the total organization. Odin Westgaard (1984) claims that his description of auditing does not imply either formative or summative evaluation. However, his ideas relate to our current description of this type of evaluation. Three types of benefits of this process pointed out by Westgaard include: (1) cost avoidance; (2) cost reduction; and (3) income enhancement. Cost avoidance is exemplified by the production of a job aid that eliminates need for training. Cost reduction is exemplified by substituting a more economical method for an inefficient one--such as eliminating an unnecessary production team. Income enhancement can occur when people become more productive than before training and produce a better product with less overhead. In general, such audits are concerned with results, not necessarily with how

results were achieved. "An operation's audit is a measure of organizational effectiveness and efficiency that can be driven by any definable operation variable, in this case, training programs" (Westgaard 1984, p. 11).

Techniques

Some of the techniques applicable here are those described by Westgaard (1984). They include measurement by (a) structured observation, (b) structured interviews, (c) survey tests, and (d) document searches. It is not likely that any carefully controlled experimental or quasi-experimental design has a great deal of relevance to studying these issues because so many variables cannot be controlled. Hamblin (1974) critiques a number of techniques that might be used for a system evaluation, and concludes that current techniques are inadequate. He suggests that such evaluations require an open-ended approach. Such an open-ended approach must inevitably concentrate less on techniques and be more tailor made and must rely more on initiative and ingenuity of managers and training specialists" (Hamblin 1974, p. 168).

Issues and Constraints

A clue to understanding the myriad of problems that may involve the assessment of training organization effects leads one to conclude that one of the most significant problems is the construction of viable techniques for the measurement of such impact. Controlled before- and after-techniques require a significant information system for the longitudinal tracking of training results. The organization must serve as its own control group in such a process, and the isolating of effects and making them directly attributable to training is difficult. Additionally, the skills and time required to apply any developed procedures may preclude their implementation.

IX. Formative Corporate Evaluation

Description

This type of formative evaluation is primarily a field-based attempt to plan for future organizational changes. Many of the processes of front-end analysis or needs assessment are applicable here. For our purposes, we will refer to front-end analysis as a general label for all activities preceding the design of a training program including an estimate of operational feasibility, projections of demand and support, needs assessment, and needs analysis.

The results of such front-end analyses may or may not have direct implications for training. The target of such an evaluation is the organization as a whole. It is similar to, but not limited to, the techniques of management auditing as described by Rothwell (1984[a]). In general, the intention is to anticipate problems and prepare an organization for dealing with change. Three alternatives for looking at this are given by Rothwell (1984[b]), Kaufman (1983) and Hunter and Cummings (1983).

In Rothwell's view, the strategic planning method (determining the long range direction of an organization) is an attempt to coordinate the activities of departments by personnel, production, finance, marketing, across several groups or industries in the same corporation. The results of a strategic needs assessment study lead to information for training as one strategy for dealing with predicted performance deficiencies. Rothwell's overall goal is stated in relation to what he sees as a significantly increasing rate of technological change and amount of information available--"We simply need to think about how to make training

capable of anticipating future performance or competency problems before they occur" (Rothwell 1984[b], p.20).

Another view of this type of "evaluation" is provided by Kaufman in what he terms "a holistic planning model." Kaufman calls this strategy an alternative to traditional needs assessment in that it is a: "holistic system approach model where we are identifying, diagnosing, and successfully treating organizational needs including determination of useful training (if training is a functional alternative)." (Kaufman 1983, p. 3).

Kaufman states that critical to this process of setting future organizational policy or directions, is the use of needs assessment, strategic planning, long range planning, and that the resulting organizational planning begin with outcomes.

A third example of such analysis is the work of Hunter and Cummings (1983). Their view of such an evaluation comprises six components:

- external environment
- internal environment
- required skills/knowledge
- existing skills/knowledge
- target audience characteristics
- non-training needs

A number of objectives are listed for each of these six functions. For example, training management is kept up to date on events from the external environment, which may influence training. Factors from this environment can be prioritized into areas of training impact. Another outcome is that management is made aware of skills and knowledge requirements to support future changes in services and products.

Techniques

Various techniques can be applied to the collection of data leading to formative corporate evaluation. In a large sense these techniques are not necessarily the domain of trainers or educators. Instead, a team is formed with backgrounds in areas such as organizational development, organizational behavior, economics, industrial psychology, labor and industrial relations, sociology, and other fields. A sampling of such techniques would include the following:

- o management audit
- o review of research on current business climate
- o review of business plans
- o analyses of planned products and services
- o longitudinal performance appraisal records
- o new hirer profiles
- o list of marketing product proposals
- o organizational analysis of all offices and departments

Issues and Constraints

Probably the most important component in linking formative evaluation to the training organization is the fact that training departments are often only minimally involved in the strategic planning process. If there is a serious gap in the planning for future human resources, it is this factor. This serves the general theme of making training reactive rather than pro-active.

X. Summative Corporate Evaluation

Description

Summative corporate evaluation by its very nature includes the complex consideration of all operating units--manufacturing, financial, personnel,

and others within an organization. This nature of evaluation requires measurement of accomplishment. Probably, the foremost model for conducting such an evaluation is the Performance Audit as proposed by Gilbert (1978, p. 61). This model implies that measurement begins at the general level and moves toward the specific. The steps of measuring follow a procedure Gilbert calls the performance table. The sequence to developing this table encompasses the following seven steps:

1. identify accomplishments
2. identify requirements
3. identify exemplary performance
4. measure exemplary performance
5. measure typical performance
6. compute the Potential for Improving Performance (PIP)
7. translate the PIP into stakes, that is, economic potential

Methods for performance auditing are different from those for management auditing as proposed in the formative model, because we are not necessarily looking at "what is", but also what should be. Both the performance audit and management audit work in tandem to identify strategic planning needs and then to see if these needs are in actuality having the desired influence in the organization.

Rothwell (1984[a]) proposes a similar model of performance audit. He defines a performance audit as "any form of research that compares a norm to an existing condition or is conducted to improve existing conditions and induce change" (p. 45). He states that performance audits assume a rational approach to organizational change, that is, justification based on the belief that decision makers will choose what is best for the organization once the facts are known. A description is supplied of a

twelve stage process to conduct performance auditing by an independent analyst reporting to an interested third party.

Techniques

Performance auditing is virtually a testing or comparison between some normative criterion of "what should be" and the condition of "what is". The two major factors that are under consideration are: (1) selection of criteria; and (2) selection of appropriate measurement methods. If the criteria originally selected are faulty, it is likely that the performance audit will not gear in on specific missions within the organization. Some methods for looking at and collecting data from a performance audit include the following:

- o information collected by industry association
- o survey research
- o cost benefit analysis
- o systems analysis
- o accounting techniques (including human resource)
- o legal information
- o information collected by government agencies
- o structured observation or interviews

Issues and Constraints

Essentially corporative summative evaluation is moving into the area of attempting to assess ultimate value of a corporation. In some cases, people limit the amount of investigation to only those concerned with financial concerns because these are the only aspects which seem to be measured accurately. This belief operates on the assumption that value means money. Hamblin's (1974) view is that the levels of evaluation proceeding to the corporate level (the training organization, curriculum,

course, and trainee) must be accomplished before we can really look at the ultimate value or goals of the corporation.

Another important aspect of this level of evaluation concerns commitment and support of line management. Support of line management and senior management is critical for performing this level of evaluation. Credibility of data depends on it. Hamblin (1974) sees three types of the ultimate value or ultimate objectives: (a) financial; (b) occupational choice--entry or opportunity; and (c) self respect.

Our experience and knowledge of this area at present is fairly limited. We are speaking from the training and development perspective, not necessarily that of a corporate executive. The techniques and methods that we use now may very well prove to be crude in the future. Nonetheless, the systems approach appears to be the best present alternative.

The expertise required to perform such an evaluation is complex, and teamwork is essential. The final decisions are really political and social more than economical.

GENERAL EVALUATION MODELS

The development of models for training evaluation has expanded rapidly over the past ten years. This is in contrast to evaluation in educational settings where growth can be traced over the past 25 years. Although this section contains references to the educational realm, the purpose for such inclusion is only to provide perspective or enhance understanding of concepts and principles.

Evaluation models presented here are approaches, systems, concepts and principles that seek a generic methodology or simplify reality to attack training issues. Evaluation studies are "designed and conducted to assist some audience to judge and improve the worth of some [instructional] object" (Stufflebeam & Webster 1980, p. 6). In comparing evaluation to research, Anderson and Ball (1978) state "research is knowledge-oriented; evaluation is decision-oriented" (p. 9) having an immediate practical payoff.

An excellent review of educational evaluation models is supplied by Stufflebeam and Webster (1980). Of the eight different approaches considered to be "true" evaluation models, most training applications would be "decision-oriented" studies where the purpose is to "provide a knowledge and value base for making and defending decisions" (Stufflebeam and Webster 1980, p. 12). Anderson and Ball (1978) reaffirm this by stating that "program evaluation involves providing services to decision makers" (p. 6).

Using the scheme of Stufflebeam and Webster, Parker (1984) classified 41 recent (1980-83) training evaluation studies and found that all reports clustered in only three categories: Decision-oriented, Objectives-based

and Accountability. Thus, training evaluation might be considered a restricted domain of the general field of program evaluation. The major question addressed by most training-oriented studies is: "How should a given enterprise be planned, executed and recycled in order to foster human growth and development at a reasonable cost?" (Stufflebeam and Webster 1980, p.18).

It is not the purpose of our review to critique training evaluation models. None of the models reviewed would be included if they hadn't worked in some context. Some models or approaches may work in a wider variety of application settings (as given in the previous section), and some were written with a single purpose or application in mind.

As opposed to models described in the educational literature, those included in the training literature are considerably shorter. Thorough understanding is often sacrificed due to publishers' constraints. Additionally, many models from corporate settings are propriety, and while we are aware of many fine examples in corporate documents, they were not included because of the limited access to our readership. Thus, we have limited our discussion to models from the public domain or those that are reasonably accessible.

A word of caution to our readers is probably relevant here. Many professional evaluators become sold on one technique or approach and attempt to apply it regardless of situational conditions. They become, in Anderson and Ball's terms, "method bound". We tend to be more problem-oriented, i.e., starting "with a premise that the choice of methodology should follow from, not precede, the delineation of the purpose of the investigation" (Anderson and Ball 1978, p. 43). Additionally, we have chosen to isolate evaluation models or techniques related to economic impact in a separate following section.

I. Model Characteristics

Each model for training evaluation has been analyzed into a number of components or characteristics described below. Subsequently, these characteristics are used as summary for Table 3 (page 42). If a given model does not appear to have a large number of characteristics checked, recall that publication constraints may have forced the authors to limit discussion.

Definition of Evaluation Perspective

Does the author describe how evaluation is to be conceived? Training evaluation is typically a team activity needing communication and collaboration among team members. How can the team remain on track? Is there a clear perspective of what is to be accomplished? How does training evaluation relate to other components in training or the organization? A useful evaluation model should allude to the answers to some of these questions.

Description of Variables/Components

Differing conceptions of training evaluation can be found to be similar if the authors defined their terms in similar language. This is generally not the case, so authors should define and describe the terms they use, because one cannot assume the intuitively obvious. To what components/variables is attention paid? Are they defined or applied to training settings? Are the key concepts described? Can variables be classified as input; process or outcome; cognitive or affective; behavior or accomplishment?

Use of Flowcharts or Descriptive Matrices

Understanding and use of an evaluation model can be enhanced by some pictorial representation. One of the outstanding features of the models

reviewed was either a flowchart or matrix of concepts or questions that guide the reader through the description. While some are more detailed than others, it is possible to obtain a simplified sketch by studying the flowchart or matrix included.

Lists of Questions to Ask

If training evaluation models are decision-oriented, then questions to direct evaluators to decision points are highly desirable. An often-mentioned step in evaluation is to define the "focus" of an evaluation effort. Focus is assisted by pinpointing issues in a question format. The authors use questions such as: Who is the client? Will management even consider this data? What should be measured? Who should be measured? How did change actually occur? What are the standards for performance? Knowing the primary set of questions a model asks increases the likelihood that it will be applied appropriately.

Data Collection Techniques

What type of measuring instrument is most appropriate to use to collect a specific type of information? Should a questionnaire be used? A rating scale? An interview schedule? Observation form? Which data collection techniques fit best with the model? Clues to answer these questions are desirable but not necessary in describing any given model.

Data Analysis Techniques

Data analysis procedures are an assumed component to evaluation, and thus not necessary to be included in the description of a model. Description of such techniques are useful in organizing an evaluation report.

Research Designs

Which research designs are most appropriate in a given evaluation? Are research designs ever really useful? Use of experimental or quasi-experimental research designs in training evaluation has been attaining greater emphasis as training evaluators become more sophisticated.

Reporting Procedures

How will the evaluation report be used? How should tables be prepared? Should interpretation guides be supplied? The effective communication of evaluation data is an often overlooked feature of evaluation.

Detailed Examples Provided

One way to get the "feeling" for how a model really works is to use a simulated situation and data. While such examples sometimes inhibit transferability, detailed explanations tend to yield increased understanding of the applications. Table 3 (page 42) is an example of such a simulation.

II. Summary of Models

It is worthwhile to begin our discussion with a model that has been used in training evaluation for approximately the last twenty-five years--the one proposed by Donald Kirkpatrick (1978). Between the time that Kirkpatrick proposed this model (about twenty-five years ago) up to

TABLE 3

CHECKLIST FOR REVIEWED MODELS

Model Authors	Def. of Eval. Persp.	Desc. of Var.	Use of Flowcharts	Lists of Quest.	Data Collect	Data Anal.	Res. Des.	Rpt. Proc.	Examples
Alden (1978)	X	-	X	-	X	-			
Bakken and Bernstein (1982)	X	X	X	X	-	-	X	-	-
Blumenfeld and Holland (1971)	X	-	X	X	-	-	X	-	-
Brandenburg (1981)	X	X	X	-	X	-	-	X	-
Brethower and Rummler (1979)	X	X	X	X	X	-	X	X	X
Clement and Aranda (1982)	X	X	X	X	-	-	-	-	-
Crowe and Bodine (1979)	X	X	X	-	X	-	-	-	-
Elsbree and Howe (1977)	X	X	X	-	X	-	-	X	X
Hunter and Nassauer (1982)	X	X	X	X	X	-	-	X	X
Kirkpatrick (1978)	-	X	-	X	-	-	-	X	-
Lyons (1972)	X	X	X	-	-	-	-	-	-
Mikesell, Wilson and Lawther (1975)	X	X	X	-	-	-	-	-	X
Morrison (1981)	X	X	X	X	X	X	-	X	-
Putnam (1980)	X	X	-	X	-	-	-	X	-
Smith (1980)	X	X	X	X	-	-	-	X	X
Snyder (1984)	X	X	X	X	X	-	-	-	-

about 1975 or 1976, there was very little other writing on models for training evaluation. Thus, this model is generally recognized as traditional in the field of training.

Four fairly straightforward steps or stages were proposed by Kirkpatrick:

1. Reaction: How do participants feel about the program?
2. Learning: What knowledge skills or attitudes were learned?
3. Behavior: To what extent did on-the-job behavior change?
4. Results: What final results did the training program produce?

These four stages of reaction, learning, behavior, and results are still present in many of the more recently proposed models. In general, Kirkpatrick was interested in answering the question, "How did we do?", by getting at the specifics of the kind of contribution training made to the organization, whether or not the program should be continued or repeated, and how it can be improved. Kirkpatrick also advocated the use of pre-/post-measures and control groups for the latter three stages of his measurement technique (i.e., learning, behavior, results).

Most present day evaluators would agree that while Kirkpatrick's model forms some basis for discussion, a more systems-oriented or comprehensive approach is needed. One such systems approach is described by Brethower and Rummler (1976) and expanded by Smith (1980).

Brethower and Rummler begin their discussion with a diagram to show how training fits with the rest of the organization. This is shown in Figure 1.

A MODEL OF A GENERAL SYSTEM APPLIED TO EDUCATION

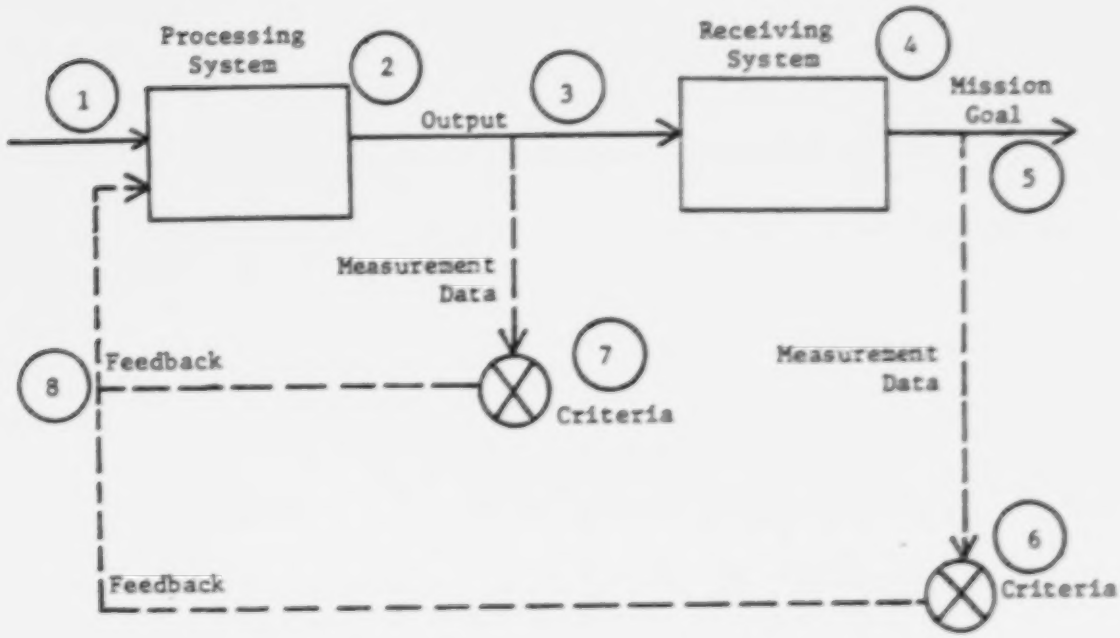

FIGURE 1

Smith enlarges this systems aproach:

> Training is viewed as a processing system (2) which converts input to output. The inputs (1) are trainees or students. The outputs (3) are trained employees whose capabilities have been enhanced through training. The outputs of training now become inputs to the receiving system. The receiving system (4) is the job or work group. The processing and receiving systems are both sub-systems of a larger system such as a company, a department, factory, or office. Each receiving system has a mission to accomplish (5) . . . points (6) and (7) represent feedback to the training administrators who can modify the course in order to increase the number of graduates who can meet end-of-course criteria or to improve the job performance of the graduates (Smith 1980, p. 70).

There are three implications for training as a result of this system's view of organizations. One is that training does not function in isolation. Second, evaluation is an information-gathering and decision making process. Third, each element in the model can be evaluated.

The next phase of the evaluation approach of Brethower and Rummler is the construction of an evaluation matrix, where the four components of Kirkpatrick's model are row labels under a column entitled "What we want to know". Five other columns are then given to explicate an evaluation design. Labels for these five columns are:

- What might be measured
- Measurement dimensions
- What to look at (sources of data)
- Alternative data gathering methodology
- Evaluation criteria

Two major benefits may result from the use of this model. First, the systems approach permits applicability across a wide variety of training situations--from looking at trainee proficiency to looking at the impact of an intervention program in an organization. It is not necessary to use all the components in each evaluation project. Rather, questions may be selected in order to obtain required information for any one evaluation setting. A second benefit is the fact that numerous evaluations designs such as control group, reversal group, multiple baseline, and before and after measures, may be overlayed into each situation.

Since Smith (1980), extends this model to a wider variety of situations, it is worthwhile to discuss his additions. Smith's contributions reflect how evaluation can be better used in a decision making process. He cites four problems that inhibit decision-making

ability from the results of an evaluation using this model. These are: no data; unreliable data that are irrelevant to the decision; untimely data; or incomplete data. He then discusses the causes of some of these problems and indicates a set of eight conditions for solutions and support of evaluation efforts in an organization. Briefly these eight conditions are:

1. Accountability - Someone above training staff requires that the information be collected.
2. Priorities - They have to be consistent with the activities conducted.
3. Decision-Orientation - Specific questions have to be addressed and can be answered yes or no in the evaluation.
4. Implicit Goals - These have to be identified. Evaluators must have knowledge of them.
5. Constraints - The amount of time available and other resources.
6. Problem Causes - Can problems in courses be detected?
7. Personnel - Qualified evaluators must be available and given proper support.
8. Viable Consequences - There is a consequence in a concrete action produced as a result of the evaluation effort.

Smith then describes two specific applications of this model illustrating the use of these techniques.

Another model falling under the systems approach is given by Mikesell, Wilson, and Lawther (1975), in which they describe an approach that integrates program development and program evaluation. They present a three-phase model of goal and feasibility determination, program design, and development implementation. While they do not state specific questions to guide the process of evaluation, they indicate a number of conditions or criteria to be followed: objectives must be established in

terms of behavioral change; training should focus on these objectives using effective forms of educational technology; trainees should be treated as adults so that their involvement is assured; careful and systematic development requires substantial front end time; lastly, the program must be constantly maintained. What is significant about their article is the description of some specific techniques for assessing the impact of the training program productivity increases in terms of dollars and cents and percent change. A description of their model shows a complete flowchart for the process as well as descriptive techniques, questionnaires, knowledge simulation and impact measures on a pre-/post-basis assessing the impact of the program.

A second type of approach to evaluation in training may be described as the management decision-making and exemplified by Alden (1978). Under this approach, upper level management must make decisions on how training affects the organization and this becomes the object for evaluation. The two types of decisions run parallel to Brethower and Rummler: Should the program change, and how should the program change? According to Alden, the most important aspect for constructing an evaluation is the focus of the effort. In order to focus evaluation properly, Alden suggests four guiding questions:

1. Will management even consider making a decision about whether or not to change a program or how it will be changed?

2. What research questions will provide data necessary to make these management decisions?

3. What level of data is practical to collect and important enough for management to use in the decision-making process?

4. What criteria will management use to make the decision?

The important point pervading these four questions is a need for commitment from management before an evaluation effort is begun. Alden sees management concern with three primary factors: effectiveness, efficiency, and relevance. Effectiveness is defined as enabling participants to perform to meet the objectives; efficiency is optimum use of program time, instructor time, materials, etc; relevance is actual performance on the job in what is critical to job success.

Alden lists a number of specific well-worded questions that would allow an evaluator to construct a fairly tight design to provide management with needed decisions. He also describes four levels of data: participant perception, expert opinion, measurement of behavior, and measurement of end results (similar to Kirkpatrick's four levels). As a final phase, he indicates how to design a reporting system in order to give management appropriate information. He advocates the use of minimum acceptable criteria so that management can get a yes or no answer to specific decisions; thus making their criteria definable:

> In this form, the question suggests a clear cut evaluation design, i.e., what should be measured, how it should be measured, who should be measured, how results should be analyzed. But most important, they offer the greatest opportunity for findings to influence the program being evaluated. Questions focused on the issues that management care about and in a form in which a no answer is a compelling reason to affect change. (Alden 1978, p. 50).

A second management orientation model is provided by Morrison (1981). Again the model is based upon management-directed questions such as:

> Who needs the information on training programs?
>
> What decisions does each individual make?

What Morrision characterizes is a set of decision-making levels where the lower level decision-making requires more specific information, and higher levels require more global information. He distinguishes four levels of decision making. Level 1 is the training staff where questions concerning the effectiveness of materials and whether participants are meeting objectives are most important. Level 2 is the training manager. Are client needs being met? Is the instructor effective? What are the costs of training? The department manager is Level 3 of this process who wants to know if the training was effective and the costs and benefits of the training. The highest is Level 4, the vice-president or director, who is interested in finding out whether or not the courses are providing an acceptable internal rate of return. Morrison presents a set of questions under each of four components to each of the four levels. The four components are: decisions to be made; the means of data collection; analysis of data; and interpretation of information. The last phase of interpretation of information is considered crucial in designing a report for all levels of decision-makers.

The next type of model to be discussed may be called a pragmatic approach to evaluation as indicated by Putnam (1980) and by Bakken and Bernstein (1982). Putnam begins by contrasting the pragmatic approach to training evaluation with the truth-seeking approach in academe. He explains how the truth-seeking approach does not work in a training setting because the pragmatic approach is future-oriented, whereas the truth-seeking approach is past-anchored. With the pragmatic approach, organizations do not have much at stake with the exact, and managers are

aware of the uncertainty that data represents truth. He presents a set of eight questions to guide the evaluation process:

1. What are the results of the evaluation intended to be used for?
2. What kinds of information count with decision makers?
3. What is to be assessed to get that kind of information?
4. What constraints on evaluation exist?
5. Whose cooperation, sanction, or approval are needed?
6. How should data be collected?
7. How should data be analyzed?
8. How shall I use the data?

He stresses the fact that evaluators should do the minimum to get it done and forget the high rigor typical of the truth-seeking paradigm.

A similar approach is presented by Bakken and Bernstein (1982) in which they advocate that the purpose of the model is to provide appropriate decision makers with the information. The general guiding questions for the authors' approach are to identify who the decision makers are, and what are the goals of training. They present a matrix of objectives of training versus the management of training function. They list four general objectives of training: personal growth; acquisition of knowledge; improvement of performance; improvement of the organization. They then list four types of outcomes: Learner Reactions, Knowledge, Job Performance, and Organizational Changes. The key in their process is to select an outcome that is appropriate to a particular objective of training and to the needs of various decision making. They also discuss the need for reliability and validity and in assessing impact to question how changes actually occurred and to assess whether changes that occurred are actually the results of training.

The next evaluation approach may be called responsive or client centered. The model presented by Elsbree and Howe in three articles (1977) demonstrates how to make evaluation responsive to the needs of the clients, that is, people making decisions about training. Their outstanding contribution is a detailed flowchart described in three phases of Focus, Plan, and Implement with a case study illustration across the three articles. Though they do not necessarily present specific questions, the steps provided in the three phases are easy to follow. In the Focus phase there are seven steps: identify purposes of evaluation; determine client information needs; determine if evaluation is warranted; find out if resources are available; define roles and job responsibilities; obtain necessary background information; and formulate evaluation objectives. Under the Plan phase, a five-step design matrix is formulated. The second of the five steps is to ask if the design will produce necessary data. Then organize tasks and schedules, estimate costs and determine if the plan is acceptable to the client. The six steps under Implementation call for: planning the execution; drawing conclusions; deciding if the information will satisfy objectives; displaying the findings; formulating recommendations; and finally, reporting to clients. The case study is easily related to actual situations.

The next evaluation model by Clement and Aranda (1982) describes a contingency approach to management training evaluation. They present four guiding questions in their design: Did it work? Was it effective? Was the manager receptive to training? Will the organization reward new behavior? The contingency approach is defined in terms of three contingency or mediating variables. The first of these is the influence

of the organizational setting; that is, finding out what effects transfer of training to the job. A second variable is the nature of the manager to be trained: what individual differences, background or experience may affect the outcome of management training? The third is the problem to be solved by the training: is the technique truly appropriate to solve the problem under consideration.

The major tool presented by Clement and Aranda is a matrix with column headings of Manager, Subordinate and Organization and four row dimensions related to purposes of evaluation: training results, the relative effectiveness of techniques, impact of individual differences, and impact of the environment. They then present a set of questions for each individual matrix cell; these allow an evaluator to collect information about the specific dimension under consideration. This model is somewhat more abstract than others given here.

A model given by Hunter and Nassauer (1982) presents in a detailed case study the results of a course development evaluation from Arthur Andersen and Company. This comprehensive, diagnostic approach is one of five phases of evaluation conducted at Arthur Andersen. Detailed descriptions of the other four phases are proprietary information to Arthur Andersen and not considered here. The five phases that are included in the total Arthur Andersen model are: front-end analysis, developmental pilot testing (which is illustrated below), testing methodology, maintenance methodology, and lastly, follow-up/performance methodology. This example only serves to illustrate one portion of the evaluation model.

The general view of evaluation taken by Hunter and Nassauer is the making of rational decisions based upon evidence. In this rational model, the following guiding questions are proposed:

- What is the business problem?
- Will timelines be met?
- What are the instructional activities?
- How many and what type of instructors are involved?
- How will evaluation information get used?
- Who receives the reports of this training? When and how?

An extensive flowchart of four phases is presented--each with input, process, and output labels. The four phases are (a) planning and preliminary design, (b) detailed evaluation design with instruments and field procedures, (c) implementation of evaluation, and (d) data analysis and reporting. The detailed case analysis presents actual instruments, simulated results, planning documents, raw data, analysis and reports all shown to illustrate their process for this one phase of evaluation.

Another type of diagnostic system is presented by Lyons (1972) which may be termed in information systems model. It is a fairly technical design where training goals must be defined precisely in terms of measurable on-the-job performance. There are four major applications envisioned by Lyons related to the general quality control of training: quality assurance, control of student progress, training program improvement, and training system diagnosis and change. Flowcharts are presented for each of these four different applications.

Under quality assurance, one is attempting to ascertain the answer to the following guiding question: "Does the product meet specifications (training materials and instructional setting which is set up to produce

the required training)?" For the second area, control of student progress, one is interested in trainee proficiency: "Are learning experiences selected and organized to facilitate achievement of objectives?" Here diagnostic results determine whether or not a trainee progresses through the system. For training program improvement, Lyons keys on the strengths and weaknesses of the program in recycling those aspects in need of revision. At the system level (the fourth application), he examines how training interacts with other parts of the organization in terms of policies, resources, and student quality. In this sense, Lyons is following a systems approach to training similar to that by Brethower and Rummler.

Snyder (1984) also presents a model with substantial illustrative flowcharts useful for evaluation of existing courses with contrasting procedures given for vendor courses. A five-stage model is described: Identification of Needs, Research Techniques (audience, task, environmental and economic/resource analyses), Instructional Design/Development, Testing, and Implementation/Maintenance. Section and combined flowcharts are displayed with decision points and pertinent questions. The purpose of the model is to list "a set of basic evaluative criteria". Detailed explanation is not supplied on all topics presented, and information relative to particular decision-makers is not discussed.

The model proposed by Blumenfeld and Holland (1971) takes a rather limited perspective of evaluation in comparison to the other models reviewed. Evaluation is viewed only as a quantitative means for demonstrating cause and effect--its only purpose is accountability. Furthermore, they see only one question worth asking: "Are we getting any return on our investment?" (p. 638). It is worthwhile to point out that

although this is their guiding question, they present no economic technique to answer the question. Instead, they advocate that evaluation consists of a simple two-step process: criterion measurement followed by experimental design. Furthermore, the only adequate experimental design consists of a pre-test and post-test with a control group. While such a design is part of the methodologies proposed by other authors (e.g., Brethower and Rummler), no other proposer has advocated it as the sole means to demonstrate training effectiveness. Readers might also consider if this approach can yield absolute cause and effect.

The last two models to be reviewed may be looked upon as measurement or empirically-based technical models. The first, presented by Crowe and Bodine (1979), applies a performance system to look at the internal and external validation of training effects on human performance subsequently translated into costs and benefits. Three very detailed documents and flowcharts on performance systems specification, training description, and measurement provide the basis for the model.

A detailed matrix is provided with questions to guide an evaluator to answer: Should you do something? Could you do something? and How you do something? The general questions include the following:

- What skills/knowledge does the student possess?
- What are student reactions to learning?
- Is there a relation between course objectives and job performance requirements?
- Is there a relation between student achievement of objectives and job performance?
- Are skills used on the job?
- Of what value is the training?

This three-phase model of Specification, Description, and Measurement is probably most useful for those who might be involved in technical skills training.

The second technical model by Brandenburg (1981) illustrates an empirical approach to looking at differences between training, education and development. Three sets of matrices are developed for general context of HRD, time, and action components. The general guiding questions are:

- What information do we need to justify a decision?
- What are the real issues or questions we wish to cover?
- What data collection techniques are congruent with the issues we have identified?

Much discussion is devoted to the different levels of decision making from specific to global similar to that proposed by Morrison, wherein needs to make specific decisions from the training staff level up to global decisions at a higher managerial level. The foremost contribution is an explanation of detailed data collection techniques consistent with particular decisions to be made in training evaluation. This model and Crowe and Bodine's are probably somewhat abstract for the typical training manager.

PRACTICE

There are, perhaps, over a thousand published reports of training evaluation studies. In contrast, there have been only a few attempts to analyze how evaluations are actually done. This section summarizes these analyses to show how training evaluations are actually done. A brief summary of the key references will be given first, and then the research on training evaluation will be reviewed in a question-and-answer format.

I. References

The earliest investigation was reported by Catalanello and Kirkpatrick (1968) who surveyed 154 firms to determine the "state of the art." Kirkpatrick (1978) repeated the survey ten years later. Campbell, Dunnette, Lawler, and Weick (1970) reviewed 73 evaluations of management training courses. Ball and Anderson (1975) reported a survey of evaluation practices in four types of organizations: Department of Defense (N=42), other federal agencies (N=33), state and local governmental agencies (N=27), and the private sector (N=40). Smith (1976, 1982[c]) reported on documented evaluations of the post-training impact of Bell System training (N=43). DeMouse and Liebowitz (1981) analyzed 36 research studies on team-building. Brandenburg (1982) described two surveys of 50 trainers on their evaluation activities. Finally, Smith (1984) described trends in training evaluation as reported in the publications of the American Society for Training and Development and of the National Society for Performance and Instruction. He catalogued 714 articles on evaluation, including 331 evaluation studies.

II. Findings

We caution the reader about several factors that limit the generalizability of the findings described below. First, the references cited above dealt with programs of differing lengths. Ball and Anderson's study had a high percentage of lengthy programs; 60% required a month or more to complete. In contrast, Smith (1982[b]) looked at evaluations of courses generally two weeks or less. Secondly, there were differences in program content. Ball and Anderson (1975) and Smith (1982) looked at technical courses primarily, while Catalanello and Kirkpatrick (1968) and Campbell et. al. (1970) considered only generic management courses. DeMeuse and Liebowitz (1981) considered organizational interventions that usually included more than formal training. Thirdly, several of the reviews examined unpublished evaluations (Smith 1982[b]; Ball and Anderson 1975; Kirkpatrick 1978; Catalanello and Kirkpatrick 1968; Brandenburg 1982), while the other studies looked only at published evaluations. These three factors--program length, program content, and published vs. unpublished--may influence the design and quality of the evaluations in ways we do not fully understand.

What "things" are evaluated?

In their survey of unpublished evaluations, Ball and Anderson (1975) found that "the focus of measurement was most frequently the students and the curriculum" (p. 18). Table 4 (page 59) shows that students were the main focus of evaluation, regardless of the purpose. In Smith's (1984) analysis of published articles, program evaluation ranked first, exceeding the total for all other categories combined.

Table 4
Object of Evaluation

Object	Ball & Anderson (1975)		Smith (1984)
	Improvement Studies	Impact Studies	Published Articles
Training program or curriculum	58%	30%	58%
Trainee or Student	63%	55%	2%
Classroom or teaching processes	46%	19%	15%
Instructors or Training Organization	34%	10%	14%
Other or Non-training program	13%	18%	8%

NOTE: Percentages exceed 100% because a study could address more than one "object".

What proportion of training programs are evaluated?

The Ball and Anderson (1975) study found that "69.0 percent of the programs had been formally evaluated..." (p. 17).

What are the purposes of evaluation?

Ball and Anderson (1975) reported that, of the 98 evaluations, 26 were concerned with improvement, 3 with impact and 69 with both improvement and impact. Brandenburg's (1982) two surveys of sales trainers and of educational technologists indicated that the two most prevalent "evaluation functions" were to "improve the training program" and to "provide feedback to program planners or management."

Who sponsors evaluations?

Ball and Anderson (1975) reported that:

> Program administration and the program's educational staff were primarily responsible for the evaluation. Outside agencies and other factors have little influence except in the case of the Department of Defense programs (p. 17-18).

Who are the evaluators?

Ball and Anderson (1975) found that the evaluations were usually carried out by "insiders," including program administrators, developers, or educational staff. In another study of unpublished studies, Smith (1982[b]) found that the program was evaluated by the same company that created the program in 33 cases versus ten cases where the evaluators came from outside the developing company. Smith's (1984) analysis of published evaluation studies found the authors almost equally divided between "academics" and "practitioners." Presumably, the academics represent mostly external evaluators and the practitioners mostly internal evaluators. DeMeuse and Liebowitz (1981) also found that the "change agent was the program evaluator."

The practical significance of using external evaluators is suggested by data from Smith's (1982[b]) report. Not only were external evaluations judged to be superior in methodology but they were far more likely to recommend major program revisions.

What criteria are used to evaluate programs?

Five reviews addressed this question, and their findings are presented in Table 5 (page 62). The studies are not directly comparable: Campbell et. al. (1970) divided criteria into only two categories, while Smith (1984) used ten categories. The major conclusion is that external measures, i.e., job behavior and organizational results, are far less used than internal criteria, such as student acceptance and learning. Unpublished evaluations (Catalanello and Kirkpatrick, 1968; Kirkpatrick, 1978) relied more heavily on reaction data while published studies (DeMeuse and Liebowitz, 1981; Smith, 1984) emphasized learning and attitude change data.

Brandenburg (1982) polled two samples of training professionals about the use of various data collection techniques (see Table 6, page 63). The two samples were in basic agreement on how frequently various techniques are used. Questionnaires and open-ended comments from students were ranked as the two most frequently used. Paper-and-pencil testing and performance testing (e.g., simulation, role play) were ranked next. It is difficult to correlate Brandenburg's findings with Smith's, but it appears that Brandenburg's respondents are more likely to use surveys and performance testing than trends in the published literature would suggest.

Table 5
Evaluation Criteria

Criteria	Campbell (1970)	Criteria	Catalanello (1968)	Kirkpatrick (1978)	DeMeuse (1981)	Criteria	Smith (1984)
Process Criteria						Content Relevance	4%
						Design Quality	7%
						Delivery Quality	3%
						Cost Efficiency	13%
Internal Outcomes	71%	Reactions	77%	75%	28%	Acceptance	37%
		Learning (Attitude Change)	50%	50%	97%	Learning	64%
						Retention	5%
						Attitude Change	8%
External Outcomes	29%	Behavior	54%	20%	39%	Post-Program Behavior	27%
		Results	10%	15%	25%	Value	11%

Table 6
Summary Of Data Collection Techniques By Criterion

Techniques	CRITERIA										
	Content Relevance	Design Quality	Delivery Quality	Student Acceptance	Learning	Retention	Attitude Change	Post-PGM Behavior	Value	Cost or Efficiency	TOTAL
Professional Judgment	2	21						1		1	25
Needs Or Task Analysis	5							2			7
Participant Opinion, Self-Report	3		4	110	9		3	27		1	157
Paper-And-Pencil Test					144	18		2	2		166
Performance Test, Job Simulation, Role Play					16			7			23
Frame Errors Or Progress Tests					28						28
Physiological Measures				1							1
Course Or Project Records (Time, Cost, Grades)				5	8		1			39	53
Instructor Rating Or Opinion		2		4	4					1	11
Classroom Observation			5		1						6
Attitude Instrument				3			23				26
Boss Opinion	1							28	1		30
Subordinate Opinion								6			6
Peer Opinion								4			4
Job Observation	1							2			3
Product Evaluation								4	1	1	5
Organizational Results, Measurements								4	31	1	36
Customer Opinion								2	2		4
TOTAL	12	23	9	123	210	18	27	89	37	43	585

Do evaluators check the reliability and validity of their measures?

Over 40% of Ball and Anderson's (1975) respondents reported that they checked both the reliability and validity of their evaluation procedures. Smith (1982[b]) found numerous threats to reliability and validity in an analysis of 23 evaluation reports, such as: known or suspected biases in the measurement procedures, unclear relationship of data to program objectives, unrepresentative sampling of graduates' work activities and vague description of data collection procedures. DeMeuse and Liebowitz (1981) concluded that "seldom was interobserver reliability or any other form of reliability or validity given" (p. 371).

For an example of an evaluation that dealt with these issues, the reader is directed to Smith (1979). This evaluation of a technical training course investigated the concurrent validity, predictive validity and reliability of eight measures of performance in training and on the job. The author also provided an analysis of biases affecting the reliability and validity of the various measures.

How large are the samples?

Smith (1980) summarized the sample sizes of 117 training evaluations. The median sample size was 55. In DeMeuse and Liebowitz's (1981) survey, "half of the studies included no more than twenty subjects."

To what extent do evaluators use experimental "controls"?

By controls, we mean such devices as control groups and pre-testing. In their 1968 survey, Catalanello and Kirkpatrick found only one use of a control group out of 110 respondents. Of the 331 studies that Smith (1984) examined, 41% employed pre-testing, 41% used a control group or an alternative instructional treatment, and only 17% used randomly selected or matched samples. Campbell et. al. (1970) found that 28 of 73 studies

used before and after measures only, while 45 studies used control groups plus before and after measures. DeMeuse and Liebowitz (1981) identified 21 of 36 studies using pre-testing, 18 studies with multiple groups. They also noted that individuals were usually not randomly assigned to alternative conditions. Of the 36, twenty studies were classed as "pre-experimental" designs which do not allow for causal inferences. Ball and Anderson (1978) also found a preponderance of one-group, "observational" studies.

How much time is needed for an evaluation?

Ball and Anderson's (1975) data show a median of two to three months. DeMeuse and Liebowitz (1981) reported a median of about six months. The latter review, it should be noted, dealt with team-building, and these interventions require more time to show measurable effects than many types of training.

Do evaluations influence decision-making?

Ball and Anderson (1975) reported the frequent claim that evaluation led to program changes (81% of the respondents). Smith (1982[b]) determined that ten of 43 studies recommended "major course revisions." The likelihood of recommending changes was linked to training received by the evaluator and whether the evaluator was employed by someone other than the company that developed the program.

ISSUES

Sooner or later, the evaluator will face difficult issues that determine the organization's support of evaluation activities and even support for the total training investment. We cannot anticipate all the critical issues that may be confronted, but our experience suggests the following represent major problems facing evaluation specialists in industry:

- o gaining the client/sponsor's commitment to act upon the evaluation findings
- o evaluating the effectiveness of the training organization itself
- o analyzing the cost/benefit impact of a program
- o evaluating management training, especially the skills that may be applied to the job in a variety of ways
- o evaluating commercially available training

Each issue will be defined. Useful approaches will be suggested. Additional references will be cited.

I. Client's Commitment

The problem is: How does one get the decision-maker to act upon the findings and recommendations? There are related questions: Who is the real decision-maker? How do you get the decision-maker to articulate the issue proposed for study? The problem of commitment can be approached from two perspectives: the immediate project at hand versus the longer term relationship with the client or decision-maker.

Our advice for a specific project expands upon planning steps 1 and 2. It is essential to know who the real decision-maker is and what the

real decision is. Sometimes, trainers will request an evaluation, but the results will be used to influence upper management or line management. If this is the situation, the evaluator must consider the information needs of the true or ultimate decision-maker. We advocate face-to-face discussion with the decision-maker to determine if the decision-maker sees the need for an evaluation study; what the decision is; what data would be considered "valid" or acceptable for enabling that decision; what the constraints are--especially time; who should participate in the planning process; and who approves the evaluation plan. Many times such interactions occur with a representative of the ultimate decision-maker. That is fine, as long as the decision-maker has, in fact, empowered the representative to act on his or her behalf. The objective is to develop a sense of "ownership" in the study by the decision-maker.

Professional evaluators often prefer to interrogate the client about the evaluation needs, go off and prepare a plan, come back and ask the client to approve the plan. That approach is often satisfactory but surprising things can happen when the planning is shared by the evaluator and the client.

A recent case involving one of the authors illustrates this point. The request was to evaluate a management training curriculum designed for one department. The evaluator proposed looking at productivity measures, comparing performance before and after managers attended training. The client did not arrange for the training to improve productivity (which, incidentally, was not a problem). Rather, the client department was experiencing high turnover in lower management levels. The client simply wanted to know if first- and second-level managers perceived any change in the management style within the department and, secondly, what needed to

happen for the style to improve. Through discussion about the department's objectives, we decided on a simple strategy.

Several groups of lower level managers were selected. Every four months, half-day discussions were held with each group about how they saw the department and higher levels of management, how useful they viewed the training, what problems they were experiencing, and what advice they had for upper management and the organization as a whole. All groups remained intact throughout a year's study. In essence, the evaluation was conceived as an ongoing "consumer research" study rather than a "one-shot" experimental study as originally proposed.

The long-term objective is to build a consulting relationship with the client organization. This should be the goal of the training organization. Evaluation should be viewed as one of the services made available to the client. One vehicle for promoting the desired relationship is a steering committee that oversees the training (or HRD) services provided to that client organization. The committee may commission needs analyses, direct the design of training programs (and other performance improvement efforts), and oversee the implementation of these programs. Evaluation becomes a means for the committee to gather data about the effects of its decisions. Thus, the client management assumes responsibility for identifying and correcting its problems, and the training or HRD organization serves as the instrument of the client.

Steering committees are typically populated with mid-level managers delegated by the senior management. The training/HRD organization is represented by someone at the same management level as the client representatives. Committees may exist for each major client departments.

Beyond the committee process, the evaluator would do well to learn as much about the client organization as possible if there is to be an enduring relationship. Some concerns include: What are the pressing issues facing the department? Who are the key decision-makers and how do they make decisions? Who are the opinion leaders in middle and lower management? How does that organization measure its success? There is no magic to discovering this information. It simply requires talking to client managers, usually in an informal fashion.

II. Evaluating the Training Organization

Top management is starting to scrutinize training operations as closely as line organizations are monitored. In part, this scrutiny is a response to a more competitive business environment. As Scherman (1980) stated: "managers faced with constricted budgets, higher costs and fewer people to do the job, are demanding a higher return on their training investment." There is a growing awareness of the huge amount of money spent annually on employee training and development. For example, Thomas (1981) stated that the annual training expenditures of GM, IBM and AT&T were $1 billion, $750 million and $750 million, respectively. Consequently, training managers need ways of demonstrating contribution to corporate goals and, secondly, efficient management practices.

We propose four approaches to measuring the training organization, its services and its internal operations. The four techniques are: (1) client satisfaction surveys; (2) focus groups; (3) operational measurements; and (4) organizational audits. All four focus on the organization, not on individual programs. Techniques (1) and (2) look at the training organization from the clients' perspective. The third

approach deals mostly with the internal functioning or efficiency of the training organization. The last is a combination of internal and external measures for a broad analysis of the training organization.

Client Satisfaction Survey

This technique involves a brief questionnaire administered by mail or telephone to managers who send people to training. Questions deal with how the client managers perceive the training organization, usually focusing on satisfaction with particular services or functions. Questions tend to be brief, and "closed-ended" questions predominate.

Table 7 (page 72) is an excerpt from a survey developed for use in a large company. It consists of 49 Likert-style questions that cover seven factors: perceived value of training, feedback from trainers to field managers about trainee performance, timeliness of training, satisfaction with vendor training, instructors' competence, client managers' knowledge of training services, and satisfaction with training facilities. This survey is used separately for each curriculum rather than for the entire training organization.

The value of this technique is that it provides a lot of data quickly about potential problems. The emphasis is on problem detection or identification rather than analysis. Consequently, subsequent data collection efforts are required to determine the causes of any problem detected by the survey.

Focus Groups

A focus group is an ad hoc group convened to discuss an issue. Focus groups have been used for "brainstorming" a problem and for market research in determining a product's image. This technique can be adapted for examining the training organization and its relationship to its

TABLE 7

Excerpt From Client Satisfaction Survey

Remember: Please read each statement carefully and circle the number which most accurately reflects your opinion.

To guide you:
"1" means the statement is true to an extremely small extent, never or not at all.
"4" means it is true to an average extent, or about normal in degree or frequency.
"7" means it is true to an extremely high extent, always or without fail.

Of course, you may use the other numbers:
"3" and "2" represent varying degress between average and extremely low.
"5" and "6" represent varying degrees between average and extremely high.

1. Course schedules are well publicized so you know when a course you need is coming up.....	1	2	3	4	5	6	7
2. Training enables my people to be self-sufficient more quickly.....	1	2	3	4	5	6	7
3. Courses are available when needed	1	2	3	4	5	6	7
4. Instructors put in a full day's work.....	1	2	3	4	5	6	7
5. Requests for training are usually met with good and timely responses.....	1	2	3	4	5	6	7
6. My training budget is a very worthwhile expenditure	1	2	3	4	5	6	7

clients. Focus groups can be organized for one occasion or they may meet periodically. Our discussion of "client's commitment" included an example of this technique.

In leading focus group meetings, it is crucial to probe for the reasons or critical events that have shaped people's impressions. Conversely, the facilitator must take care not to bias opinions by inadvertently encouraging or discouraging certain types of comments.

Operational Measurements

An organizational measurement is a statistic which describes some characteristic of an organization's internal operations or external services. A measurement is timebound, meaning that it portrays the organization for a specific period of time, e.g., monthly, quarterly. Measurements may deal with such training functions as program development, delivery, and scheduling. Measurements may represent such dimensions as volume (frequency, amount), cost, efficiency, timeliness or quality. Examples include: average cost to train one employee for one day, "seats" scheduled as a percentage of "seats" requested, percent of students successfully completing training. Training measurements serve as problem indicators. They are a "warning system" which should trigger further data collection to determine the causes of negative trends.

Table 8 (page 74) shows a 1984 survey of the training measurements used by five former Bell System companies. Many of these statistics were experimental and are no longer used. This list is presented to illustrate the variety of possibilities.

There are two common problems with using measurements to evaluate training organizations. The first problem is the unfair comparison of training groups. Training measurements are affected by such variables

Table 8

Partial Listing of Training Measurements Tried By Five Companies

MEASUREMENT	Co. "A"	Co. "B"	Co. "C"	Co. "D"	Co. "E"
Relevance (rated by graduates)	X	X			
Effectiveness (rated by supervisors)	X	X			
Delivery cost per trainee (on hour)	X	X	X		
Demand vs. capacity in terms of student days	X		X	X	
% seats cancelled (management and non-management)	X		X	X	
% courses developed @ Training Development Standards	X	X			
% projects completed on time	X	X			
% seats provided to requested seats		X		X	
% Training Manager observations completed		X			
Average time in initial training		X			
Student days per instructor day	X	X	X	X	
Student/instructor ratio	X	X			
% follow up evaluation completed/committed		X			
Developer hours/instructional hour	X	X		X	
Development cost/instructional hour	X	X			
% developer days spent on developing		X			
Trainee cost/trainee hour by training center			X		
Instructor utilization (% time in class)	X				
No. trainers per 1000 employees					X
Delivery hours per 1000 employees					X
Delivery hours/1000 customer contacts					X
Delivery hours/number of new hires, transfers, promotions					X
Delivery hours/$1000 capital investment					X
Delivery hours/$1000 revenue					X

as: differences in client demands, e.g., seasonal fluctuations, or high volume courses versus low demand courses; differences in delivery strategies, e.g., self-paced versus group-based training, or field site versus training center; differences in program development, e.g., purchase of "off-the-shelf" courses versus contracting for customized courses versus "in-house" designed courses; differences in staffing strategies, e.g., part-time versus full-time trainers, use of contractors, use of instructors borrowed from other departments. Training measurements should be used to understand critical variables that shape the training group's activities. Appropriate standards come from historical trends developed for a particular training group rather than inter-group differences.

The second problem is the use of self-reported measures which are easily distorted. We are familiar with one training department that relied upon project managers to report project completion dates versus project target dates. Over a four year period, no project was ever completed behind schedule. The most reliable measurements are often derived from data gathered for other purposes or by non-trainers. For example, the "cost per trainee hour" can be compiled from the course scheduling process (trainee hours) and the accounting process (cost data).

Organizational Audits

An audit is an intensive study of an organization for the purpose of finding ways to improve its operation. Individual programs are evaluated but only as samples of the organization's work. Audits may employ a variety of data collection techniques and usually require a substantial investment of time and people. Audits are usually staffed by people from outside the audited organization.

Table 9 (page 77) comes from one company's audit plan (Smith, 1979). The columns represent methods for collecting data. The rows represent variables which can be measured in an audit. Other audit plans have been presented by Tracey (1968), Lien (1979), Gaskell and Svenson (1978), Chellino, Rice and Dinneen (1978) and Olivas (1980).

Audits generate much information but at great cost. It is important to involve the audited management team in defining study objectives and, secondly, in interpreting data. The goal is to develop the commitment of training managers to act upon the audit findings.

III. Cost Benefit Anaylses

Deciphering Training Cost Data

The prerequisite for understanding cost analysis in training evaluation, whether it is cost benefit analysis, cost effectiveness analysis, or other types of evaluation, is to look at costs themselves. As an initial phase, we will examine the cost of training input only, excluding for the time being those costs associated with benefits or other outcomes. According to the economic definition, cost is the sacrifice of an alternative. For example, one training program is put on in deference to another. The general problem with costs is that they are thought of too narrowly, more or less in terms of a line item in a budget. These may be called accounting costs typically obtained from the accounting department. It is also likely that such costs are a distortion of the true tangible and intangible costs observed in the design, development and delivery of a training program.

It would seem that the measurement of costs would be relatively simple if all we would have to do is scrutinize accounting statements and select

TABLE 9

Training Evaluation Exercises and Variables

Variables	EXERCISES																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
	Trainee Questionnaire	Trng. Delivery Mgr. Interview	Analysis of Seat Cancellations	Analysis of Classroom Use	Analysis of Instructor Use	Instructor Interview	Analysis of Test Procedures	Analysis of Test Data	Observation of Instructors	Graduate Questionnaire	Methods Mgr. Interview	Review of Course Content	Anal. of Dvlpmt. Project Statistics	Compliance with Development Stds.	Trng. Development Mgr. Interview	Graduate's Supervisor Questionnaire	Observations of Graduates	Field Manager Interview	Anal. of Delivery Expenses	Anal. of Instructional Design	Compliance with Adaptation GL
Scheduling Process		X																X			
Timeliness of Training	X																				
Prerequisites (Compliance)	X	X				X															
Pre-training Briefing of Trainees	X	X				X															
Training Requests vs. Scheduled Training		X																X			
Seat Cancellations		X	X																		
Curriculum Planning		X																			
Classroom Utilization				X																	
Instructor Utilization					X																
Instructor Performance (Evaluation)								X	X	X											
Procedures for Evaluating Instructors		X				X															
Testing and Grading						X	X	X													
Delivery Expenses																			X		
Learning Problems of Trainees						X		X													
Post-Training Work Assignments		X								X						X					
Post-Training Job Proficiency										X						X	X				
Training vs. Field		X														X		X			
Feedback from Field to Trainers		X																X			
Instructors vs. Developers		X													X						
Development Priorities		X													X						
Timeliness of Development		X											X		X						
Course Implementation Problems		X				X									X						
Instructor's Change of Course						X									X						
Productivity of Development Group													X		X						
Efficiency of Development													X								
Efficiency of Adaptation													X								
Cost of Development, Adaptation													X								
Compliance with Training Dev. Stds.														X							
Compliance with Adaptation GL																					X
Quality of Materials & Design		X				X	X			X				X						X	
Relevance of Course Content						X				X		X				X					
Evaluation of Development Projects															X						
Methods Staff vs. Instructors		X									X										
Methods Staff vs. Developers															X						
Conversion Training		X									X				X						
Special expertise required?	No	No	No	No	No	No	Y-	No	Y-	No	No	Y-	No	Y-	No	No	Y-	No	No	Y-	Y-
Preparation Time (in days)	1	½	¼	¼	¼	1	½	½	1½	5	½	½	½	½	½	5	1½	1½	½	1	½
Administration Time (in days)	3	½	0	0	0	1½	0	0	3	15	½	3	0	0	½	15	8	1½	0	6	0
Analysis Time (in days)	1	½	¼	¼	¼	1½	1½	1½	1½	3	½	1½	3	3	½	3	3	1½	1	3	1½
Sample Size (*per course)	10	1	NA	NA	NA	1*	NA	NA	1*	15*	NA	NA	NA	NA	1	15	5*	3	NA	NA	3

BEST COPY AVAILABLE

appropriate numbers. Most authors reviewed state that this is almost never the case. In fact, in comparison to other variables designed and measured in a given evaluation, it is likely that cost data would tend to be crude and incomplete by comparison. For example, sometimes it is difficult to separate the developmental costs of a program from the operating costs or administrative costs, or from pilot testing. The major reason why the cost estimates tend to be incomplete is that they are based upon actual expenditures rather than the costs of all resources. What is probably called for is some sort of cost information system geared to the actual training program design, development and delivery. (See Smith 1982[c], for an example of such a system).

For our purpose, we reviewed five different attempts to categorize cost data displayed in Table 10 (page 79). The first scheme of Cullen, et. al. (1978), is from an experiment run on production personnel in an actual training experiment. This list is not as extensive as other lists even though the resulting information was used in a cost effectiveness study. The second data set from Mirabal (1978) and Deming (1979) (U.S. Civil Service Commission) offers five basic documents for cost categorization. Four basic components are those given under Step 5: Course Development, Participant, Instructor, and Facilities costs. Using all data, one is able to break down total costs as indicated under number 6 according to four categories--on an annual basis, per trainee hour, per curriculum hour, and per trainee.

A third set of costs is given by Spencer (1984). Claiming that direct labor of training personnel is grossly underestimated, Spencer estimates fringe benefits and overhead, and states that the total labor involved in training staff is about three times the direct labor involved. The fourth

Table 10
CATEGORIES OF COST DATA

Cullen, Sawzin, Sisson & Swanson (1978)	Mirabal (1978) Demming (1979)	Spencer (1984)	Weinstein (1982)	Smith & Marcinuk (1982)
1. Training Development analysis time design time material costs 2. Training Materials (expendable) cost of reproduction 3. Training Materials (nonexpendable) instructional hardware instructional software 4. Training Time trainee time trainer time 5. Production Loss Resulting From Training production site losses material losses	1. Trainee Costs salary travel per diem material & supplies tuition 2. Instructor Costs hours salary (preparation & instruction) travel per diem 3. Facilities (non-organizational) rental improvement equipment 4. Development salary (in-house) contract fees production costs travel per diem 5. Obtain Specific Costs From • Course Development Worksheet • Participant Cost Worksheet • Facilities Cost Worksheet 6. Get Total Costs • annual basis • per trainee hour • per curriculum hour • per trainee	1. Development of Course direct labor total labor (above x 3) materials expense 2. Trainer Training direct labor total labor (d1 x 3) per diem materials travel facility cost 3. Delivery of Training trainer labor materials facilities 4. Evaluation/Follow-Up evaluation labor expenses	1. Classroom Expense outside instructor in-house instructor fringe benefits travel per diem materials classroom use visual, mechanical aids food, refreshments other 2. Administration Expense administrators & fringe typing, clerical curriculum development postage, shipping, telephone program, printed materials other 3. General Organizational Costs general organization support top management (salary & fringe prorated) 4. Participating Compensation (opportunity costs)	1. Curriculum Development Costs salaries benefits vouched expenses vendor program purchase salary & expenses of staff supporting effort (e.g., SME) 2. Delivery Staff Costs salaries benefits vouchered expenses vendor program purchase license fees, royalties materials salary & expenses of supporting staff 3. Facilities Cost leases, taxes, utilities, insurance maintenance, depreciation participant expenses 4. Administrative/Support Costs salaries for support groups (e.g., clerks) salaries for administrative groups benefits & expenses apportioned salaries & expenses of senior management apportioned overhead for corporate service 5. Client Organization Costs trainee salaries & benefits vouchered expenses tuition reimbursement tuition for vended programs temporary hire lost productivity apportioned salary & expenses of line training coordinator apportioned salary of supervisory time for enrolled employees

BEST COPY AVAILABLE

model is from Weinstein (1982), who categorizes three general levels of costs--those associated with classroom, administration, and general organizational costs. Note that he advocates the inclusion of costs where top management must take some interest in the training program. Additionally, opportunity costs are included as a separate item.

The fifth attempt (Smith and Marcinuk 1982) is a compilation of lists acquired across a number of sources as well as an internal investigation by New England Telephone. Although Smith claims that the list is not exhaustive, it is certainly more comprehensive than others reviewed. Opportunity costs, line and senior management costs and replacement of trainee costs are among the diverse categories reported. The results of the calculations proposed by this scheme feed into the information system described by Smith elsewhere (Smith 1982[c]).

The purpose in presenting these different costs categorizations is to allow the reader to judge the completeness and comprehensiveness of cost schemes proposed by others. In general, costs associated with personnel are generally in the range of 70 to 80 percent of any training budget. One guideline from Levin (1975) is to spend no more time on investigating a cost than its contribution to a budget. Thus, if a given item is around two percent of the budget, then one should spend no more than two percent of the time trying to get actual costs.

Most authors reviewed agree that the analysis of training costs is very difficult and that calculating something like "return on training investment" is almost impossible. Then, why even attempt it? What are the benefits of trying to track costs? Smith and Marcinuk (1982) provide some solid reasons for such undertakings. Probably most important, top management scrutinize budgets for reallocations and demand higher return

for their training investment. A second reason is that accounting for training will assist in operationally defining what training actually means within the organization. A framework for deciding what may or may not be training is given by Smith and Marcinuk (1982). Thirdly, training managers might use such cost data for detection and analysis of inefficient practices, identification of key factors which influence costs, projection of training budget, evaluation of programs or service, and propose actions in terms of impact on costs. Finally, the clients of training, mainly line management, need to be convinced of the value of training. Most often, it is their budget and operating results to which training is charged.

Analysis Models

Parallel to the increased development of training evaluation literature in general has come a fairly significant increase in writing devoted to economic analysis for training. These analyses range from the simpler techniques such as return on investment and payback to cost effectiveness, and cost benefit analysis. The purpose of this section will be to describe, but not in detail, a number of models that have been developed to look at the impact of training from a financial or at least a quantitative perspective.

Most authors reviewed comment about the lack of knowledge that training personnel seem to have with the economic side of their enterprise. Prior to the 1980s it is fairly apparent that there has been little use of economic analysis methods in training evaluation. Thus, the number of detailed examples is small. Additionally, most training managers and evaluators have not become familiar with cost analysis approaches to decision making. In the field of training, rarely do cost

considerations other than keeping within budget and time constraints. really control the training process. Thirdly, there are few textbooks in training area that can provide satisfactory introductions in how to do appropriate cost analysis, and many trade journal articles result in confusing readers with inaccuracies.

Presentation of information for this section is facilitated by the Appendix (page 105) which lists nine models that have been discussed in the training literature. Each model is outlined in terms of six descriptions: proposer/proponent, purpose, use, basic inputs/assumptions, formula/method, outcome guidelines. The general progression in reviewing these models is from the simple to the complex.

The first two models--Return on Investment (ROI) and Payback Period--are fairly simplistic approaches to looking at economic impact. It is assumed, in order to perform these analyses, that one needs a fairly accurate accounting of costs. What is generally not taken into account in ROI analysis is the use of present values. In simplistic approaches to ROI, the present value of money is not considered. Barta (1982), however, advocates that this can be a fairly rigorous approach if proper procedures are followed.

The first four models, ROI, Payback, Human Capital and Measurement-Based analysis models are the only approaches reviewed used for examining a single training program or training project. It is necessary in cost-effectiveness, cost-benefit and other analyses to make a comparison among two or more alternatives.

Odiorne's (1979) Human Capital model consists of a conceptual rather than a pragmatic approach of looking at training contributions. An economic analysis can tell whether or not to train at all according to

Odiorne. He indicates that the goal of training is not behavior improvement, but profit improvement for the organization. In order for training to make a contribution, the economic objective of training must be satisfied. If it doesn't contribute, then it should be eliminated. One of Odiorne's unique contributions is to classify economic objectives of training with regard to immediate, intermediate and long-range, with the long-range objective being the Human Capital portion.

Measurement-Based analysis (Nickols, 1979) is not so much an evaluation model, but in fact the opposite. Nickol's approach is to examine any quantitative based system in order to connect organizational means (activities) to ends. The purpose of this analysis is to define whether or not an intervention (training or otherwise) should occur. Thus it precedes, not follows, evaluation. Because his model is economic in nature, it is worthwhile including it here as a potential alternative. One of the important outcomes of this analysis is to quantify the cost of a problem in relation to organizational standards and subsequently to quantify the value of the solution. In this way, Nickols is able to separate environmental effects from individual behavior.

The fifth model is the Life Cycle approach given by Kearsley and Compton (1981). One of the major purposes of the Life Cycle model is to recognize that training occurs in a cycle; that is, there is a start-up, an operational, and then a transition phase to any new activity. During the transition phase, there are costs associated with at least two programs. The major point of this analysis is that costs of two or more training approaches must be compared at the same point within their life cycles in order to make a fair comparison.

The sixth model is Productivity Function analysis, also described by Kearsley and Compton (1981). To the reviewer's knowledge, this is a theoretical model which has its roots in classical economic theory and has not been applied to any training settings. Productivity analysis uses a multiplicative function to, in one case, predict amounts of money that can be spent on a given training activity. One of the major outcomes is a set or series of productivity curves which may be compared across alternatives with steeper curves showing more productivity or outcome. Probably one of the most practical implications of this model is the recognition that there is always some point in any given training approach where limitations are reached. That is, further increases in, for example, the amount of time allotted a trainee to perform at a given level, can only be increased so much until the costs greatly outweigh potential accomplishments of time increases. The primary use of this model is for large training organizations and sizable projects.

The seventh model is the Economic Benefits Forecasting model or Performance Value analysis proposed by Swanson and Geroy (1984). This is a simplified version of cost-benefit analysis. The purpose is to forecast training benefits accrued by two or more programs to facilitate comparison and thus, the decision-maker's choice among alternatives. Essentially the model consists of three stages: (1) listing all training costs specifying the total performance in terms of dollars and cents; (2) assessing or judging the value of a given performance unit; and (3) subtracting the cost from performance value yielding net benefits. The approach with the largest net benefit is then the choice of the decision maker. Swanson and Geroy provide two worksheets and a companion application article to illustrate the steps of their analysis.

Cost-Effectiveness and Cost-Benefit analysis are probably some of the least understood and most misapplied tools in training and educational evaluation. Different authors seem to define their terms differently. In both cases, the goal of either analysis is to select an alternative from a number of alternatives, in our case the most beneficial or effective training strategy. In Cost-Effectiveness analysis, we are looking at the cost of two or more activities designed to produce very much the same outcome. In Cost-Benefit analysis, common objectives among programs is not a prerequisite. Thus, one could do Cost-Benefit analysis to decide between management training and technical skills training. With Cost-Effectiveness, we would only be concerned with skills training for a particular type of person doing a particular type of job. In Cost-Effectiveness, as its name implies, we use an effectiveness index, which is generally some measurement of whether or not one or more objectives of the program have been met. In Cost-Benefit analysis, benefits are extracted and values are assigned to these benefits. Cost-Effectiveness analysis can account for tangible and intangible effects. In Cost-Benefit analysis, however, one must convert both tangible and intangible benefits to monetary terms. This is sometimes criticized as one of the most important shortcomings of Cost-Benefit analysis--that is, the quantification of benefits. Additionally, there is the problem of precise identification of those benefits.

In sophisticated explanations of Cost-Effectiveness and Cost-Benefit analyses, present value terminology, including a discount factor, is used in order to provide more precision of dollar values. The choice of a discount factor is another point of contention for such analyses, because of its sensitivity to results. Most authors would seem to agree that the

use of Cost-Benefit analysis in training evaluation probably requires some stretching of reality. Those who advocate use of this approach would like, at a minimum, to see it used as a screening mechanism or forecasting tool such as proposed by Swanson and Geroy.

Cost-Effectiveness analysis requires programs having identical or similar goals. Common measures of effectiveness must be used or interpretation is impossible. Cost-Effectiveness does not lead one to conclude or establish whether or not a training investment is worthwhile--only the potential cost and the relative effectiveness among techniques.

IV. Evaluating Management Training

The evaluation of management training is a hotly debated topic. Some commentators feel that management training cannot or should not be evaluated because the cost of the evaluation cannot be justified, the effects of management training are too subtle to be reduced to numbers, and intervening factors obscure any cause-effect relationships. A second question is how to measure management training and involves deciding between opinion measures (either self-reported or boss-reported) versus more sophisticated procedures. The latter would be typified by quasi-experimental designs using job behavior or job output measures.

In this section, we will try to clarify the problems associated with evaluating management training and offer suggestions for how to cope with these problems.

The first issue is: What is management training? At the risk of over-simplification, let us define three types of programs that may be called "management" training:

1. Specialist or individual contributor courses, i.e., non-supervisor work assigned to management people, for example, selling and engineering tasks;

2. Self-management courses, e.g., time-management, certain kinds of problem-solving and planning courses;

3. Inter-personal training for managing other people or getting work done with peer work groups, e.g., leadership courses, group problem-solving, a course in how to give performance feedback to a subordinate.

While all three types involve measurement problems, it is the last category that commentators generally have in mind when they discuss the problems of assessing management training.

There are several problems in measuring the effects of this type of training. First, using the newly learned skill may require a particular situation, such as an attendance problem. This situation may not be predictable or may occur some time after training. Therefore, it may not be convenient to record the application of the skills, or the proficiency level may deteriorate through forgetting or disuse. Second, the new skills may relate to a variety of situations as, for example, skill in how to analyze work performance problems. The manner in which the skill will be applied, and consequently, how its application can be measured, may not be predictable. A third measurement problem is that interpersonal and managerial skills may not be directly observable, either because the skills are covert or are exercised in confidential situations. Finally, the desired effects of these managerial skills may be defined in terms of other people's behavior, for example, subordinate's productivity.

Measuring behavioral change in other people often takes a long time because the desired changes result from the accumulated effects of many interactions in which the managerial skills are used. The longer the time to measure the effects of training, the greater the odds for other factors to influence job behavior and obscure the effects of training.

There is a second set of problems in evaluating management training in terms of its effects upon job performance. No matter how good the training, other factors may inhibit or discourage trainees from using what they have learned. The most important influence on a manager is the boss. If the boss does not encourage or reward the trainee for applying the skills learned in training, there is a good chance that the skills will not be used. More subtly, the boss's behavior represents a model for the trainee and may be more effective than any intentional act to shape the trainee's behavior.

Beyond the boss, there may be organizational factors that support or discourage the trained skills. Some of these factors may be features of such company systems as measurement systems (productivity, quality control, sales), compensation programs, performance appraisal, promotion criteria, work flow structure. For example, if the training deals with team-building, but the measurement systems encourage competition among work groups, then the training will probably have little effect.

Besides the formal systems, there may be significant informal systems that may inhibit change. These systems are generally social or communications systems. Examples include: cliques within the management hierarchy who may be competing for power, the trainee's peer group who may prefer a status quo, employee groups who may be antagonistic to management.

Our task here is not to overcome these "support" issues but how to evaluate training given these obstacles. Nevertheless, we need to say something about appropriate training/change strategy as a basis for evaluation strategy. Management training is more likely to have measurable benefits if the following tactics are used:

1. Do not use management training by itself to accomplish important organizational changes, such as changing a department's management style.

2. Imbed training in a broader strategy that may include (a) changing organizational systems to be compatible and reinforce the change objectives, (b) developing the active support of higher management in such areas as informal rewards, performance feedback, leading by example, (c) developing employee awareness and support.

3. Deliver the training in a style that supports the change objectives. If team-building (or "team-playing" as one colleague prefers to call cooperation between permanent groups) is the goal, then sessions should be populated with the people who are supposed to work together. If the skills have long learning curves, then follow-up sessions may be worthwhile. Sometimes, these follow-ups are devoted to experience-sharing, problem discussion and action planning.

4. Consider separate training for higher-level managers on how to reinforce the behavior of lower level managers. Often, this training aims to bring consensus among the higher levels on change goals and strategies.

5. Allow substantial time for implementing the change strategy. A rule of thumb is two to five years for an organizational change to be successful.

These points lead to several inferences about appropriate evaluation strategy. First, the entire change strategy should be the target of evaluation, not just the training. Conversely, if there is no comprehensive strategy, then evaluation efforts cannot be expected to detect any measurable payoff and may not be worth the cost. It also follows that organizational measures, not measures of the individual trainee, should be the basis of evaluation. Such measures may include:

productivity, turnover, absenteeism, grievances, morale surveys, work quality. The evaluator should look for baseline measurements which can be used for "before-after" or inter-groups comparisons. Quasi-experimental designs have been used effectively in these situations. The reader should consult Brethower and Rummler (1976, 1979), Cook and Campbell (1976) and Smith (1982[a]) for useful discussions of evaluation designs.

Second, the evaluation plan should assess each part of the change strategy in terms of whether the change tactics were implemented as designed and what problems were encountered. One approach is to outline the project in terms of milestone events and results expected. Tracking the project in terms of these milestones provides a basis for corrections. This approach is often dictated by the lengthy time to accomplish an organizational change.

Another approach is to use focus groups periodically for getting participant opinions about the progress of change.

V. Evaluating Commercially Available Training

In large numbers of organizations, training materials are not developed in-house. Instead, all courses and seminars are purchased from external vendors. In many other organizations, internally developed training is supplemented by vendor courses. In these two primary cases, evaluation of vendor courses is an important step in selecting activities to meet training needs. Our purpose is to assist the selector in choosing among two types of vendor training: seminars and packaged programs. Initial discussion is devoted to seminars and workshops.

There is no doubt, from reading trade magazines, that the seminar business is big business. Cost for attendance at a seminar per employee

for a one-week program could run in the neighborhood of $1500 to $3000, depending on the type of employee, content and contributions from the employee organization. Is a business likely to spend $3000 on equipment without some prior information for comparison of different brands and models? Probably not. Why should selecting seminars be any different? How do we organize a process to sift through these materials?

One procedure developed by Martin (1983) provides a workshop buyer's checklist divided into seven components. The seven components are:

- o needs assessment
- o workshop objectives
- o workshop content
- o workshop design
- o environmental issues
- o workshop leadership style
- o evaluation.

The complete 39-item checklist provides careful, initial examination of training seminar brochures. A common next stage of the analysis is to gather additional information as suggested by Phillips (1983): being wary of the brochure; gathering additional details; checking out the speaker; talking to previous participants; investigating the learning environment; and securing additional evaluation data.

Phillips completes his analysis by four additional steps. The first of these is Pre-Programmed Planning whereby individuals in the training organization contact line management to support an investigation of a given seminar. Once seminars are selected for tryout, an evaluation form is sent to pilot participants prior to seminar attendance. The next stage is Post-Program Activity. Participants are followed up through interviews

or letters requesting what value they found by attendance at the seminar. The last two steps are: Building a Data Base, a centralized function to accumulate data from pilot attendance, and finally, Follow-Up, three to six months after the seminar with a questionnaire and letter to further investigate benefits of the seminar. In summary, this common sense approach to evaluation, following the above steps, should increase the value of seminar or workshop attendance.

A second, more complex process, is the evaluation of vendor packaged courses. These are distinguished from seminars in the sense that they may or may not be offered with a standard speaker and are often smartly packaged materials available for purchase by the organization to be used in-house.

Let us first examine some of the reasons for why it may be necessary to develop an evaluation guide for such courses:

- Training departments are often given budget to buy, not necessarily develop, a certain number of hours of training.
- Millions of dollars of package training are already available for a wide range of courses that could meet organizational needs.
- Many training packages have highly developed marketing strategies and slick promotion often makes it difficult to judge the quality of any given package.
- True customization or the offer of customization is often an attractive sales pitch.
- Evaluating the potential use of a training package may require thorough piloting costing $10,000 to $20,000.

In reviewing available literature on evaluation of vendor courses, it is obvious that most authors would agree on one primary conclusion--do everything possible to stay away from a pilot or field test until all other criteria are met. The reason is that the pilot test is the most

expensive, as well as the most vulnerable, step in this process. (Vulnerability is experienced by any training unit sponsoring a "bomb").

The trends in purchasing of packaged programs, according to recent surveys of training, is that there are going to be more available, and there is likely to be more customization. Additionally, these surveys point out that the primary criteria most organizations use to select package programs include: the ability to customize programs; cost; and previous experience with the supplier.

Our review indicated a fairly uniform set of stages to follow in the evaluation of vendor courses. Each of these stages involves successive levels of commitment to the program. In most cases, these stages can be utilized to evaluate a single course or to compare a number of alternative courses. Three primary references used in this analysis include Phillips (1983), Rogers and Volpe (1984), and Ross (No Date).

An initial stage suggested by Rogers and Volpe is that of "weeding out junk mail." This stage consists of a few simple questions that may be undertaken by the training staff, including: Are there stated learning objectives? Is training performance-based? Is there a topical outline? Does it appear to have more substance than simply a slick marketing attempt?

The second stage in this process consists of a short checklist like that suggested by Phillips (1983), where a series of 15 evaluation questions are used to determine the appropriate criteria and potential match of the content to the needs of the organization. It is a preliminary stage of analysis, not necessarily including the use of subject matter experts. Rogers and Volpe attempt to sample materials, including the content outline, find out what participant materials are like, obtain testimonials from other organizations, and estimate costs.

The next stage in successive examination includes the building of a content review or a matrix of course objectives applied against principal job outcomes. Here it is desirable to have subject matter experts rate how well a course simulates the job requirements (Ross 1977). An alternative procedure, although very similar, is a ranking of proposed programs in terms of criteria and importance performed by both training and line management. A matrix for this process is provided by Rogers and Volpe (1984), listing the tasks and validity ratings across a number of criteria for each vendor course. For Rogers and Volpe, a subject matter expert is used in a further stage called "instructional design analysis" where an instructional designer carefully reviews such topics as match of objectives and content with needs analysis, internal consistency among instructional objectives, identification of prerequisite skills, sequence of instruction, appropriateness and adequacy of instructional strategies and completeness.

The next stage is an internal review of the potential cost benefits for purchasing and administering the program within the organization. This level of analysis suggested by Ross takes into account anticipated payoff in relation to purchase price, trainee wages and expenses, and opportunity costs. Such an analysis also considers the "value of task performance" associated with proposed course outcomes.

The final stage, field or pilot testing, is the most expensive step of analysis. Most courses will have been eliminated prior to this stage. In Ross's terms, there are three important pieces of information to gather: learner reactions, internal validity information, and external validity information. Learner reactions can be gathered through the use of a standard questionnaire or follow-up interviews with participants. (Rogers

and Volpe supply a sample questionnaire used in their analysis.) Internal validity includes data collected during the course which has potential for follow-up later. External evaluation is the typical summative information collected for follow-up on a three to six month basis after the course is delivered to the pilot audience.

A final question is: What happens after all of these stages are completed? Results found by Rogers and Volpe included the following: an instant drop in the vendor's price; vendor interest in the report to share with the home organization; a data base for training needs analysis; a valuable down-to-earth report in non-training language; a matrix to be used for further evaluation; and a strong critical capability within the organization to negotiate with a vendor in a clear, forthright manner. A final point raised by Phillips is to take advantage of "guaranteed results." If the supplier uses this tactic, it permits the training organization a potentially valuable further recourse, especially if large expenditures are involved.

REFERENCES

Alden, Jay. "Evaluation in Focus." Training and Development Journal 32 (October 1978): 46-50.

Alden, Jay. "Troubleshooting Evaluation." Paper presented at the annual meeting of the National Society for Performance and Instruction. Dearborn, Michigan, April 1983.

Allen, D.C., H. Amacher, and J.P. Yaney. "The Economics of Management Change; Cost Estimating." Paper presented at the annual meeting of the National Society for Performance and Instruction, Atlanta, Georgia, April 1976.

Anderson, Scarvia B., and Samuel Ball. The Profession and Practice of Program Evaluation. San Francisco: Josey-Bass Publishers, 1978.

Bakken, David, and Alan L. Bernstein. "A Systematic Approach to Evaluation." Training and Development Journal 36 (August 1982): 44-51.

Ball, S., and S. Anderson. Practices in Program Evaluation: A Survey and Some Case Studies. Princeton, N.J.: Educational Testing Service, 1975.

Barta, Tom. "Methods to Determine Return on the Training Instrument." NSPI Journal 21 (February 1982): 16-17, 24.

Blumenfeld, Warren S., and Max G. Holland. "A Model for the Empirical Evaluation of Training Effectiveness." Personnel Journal 50 (August 1971): 637-40.

Brandenburg, D. "Training Evaluation: What's the Current Status?" Training and Development Journal 36 (August 1982): 14-19.

Brandenburg, Dale C. "Quality Assurance of Human Resources Development Programs: A Scio-Technical Continuum Approach." Paper presented at the annual meeting of the American Educational Research Association, SIG/Training in Business and Industry Symposium, Los Angeles, California, April 1981.

Brenden, Don. "Target Evaluation System." Paper presented at the annual meeting of the American Educational Research Association, SIG/Training in Business and Industry Symposium, Los Angeles, California, April 1981.

Brethower, Karen S., and Geary A. Rummler. "Evaluating Training." Training and Development Journal 33 (May 1979): 14-22.

Campbell, J.P., M.D. Dunnette, E.E. Lawler III, and K.E. Weick, Jr. Managerial Behavior, Performance and Effectiveness. New York: McGraw-Hill, 1970.

Catalanello, R.E., and D.L. Kirkpatrick. "Evaluating Training Programs - the State of the Art." Training and Development Journal 22 (May 1968): 2-9.

Chellino, S.N., R.L. Rice, and M.A. Dinneen. "Corporate Training Audit." Paper presented at the annual meeting of the National Society for Performance and Instruction, San Francisco, March 1978.

Clement, Ronald W., and Eileen K. Aranda. "Evaluating Management Training: A Contingency Approach." Training and Development Journal 36 (August 1982): 39-43.

Cook, T.D., and D.Y. Campbell. "The Design and Conduct of Quasi-experiments and True Experiments in Field Settings." In Handbook of Industrial and Organizational Psychology, edited by M.D. Dunnette, 223-326. Chicago: Rand McNally, 1976.

Crowe, Martin H., and Robert L. Bodine. "The Three Phases of Evaluation: Specification, Description and Measurement." Performance Systems Group, Inc., 1979. Workshop materials.

Cullen, James G., Stephan A. Sawzin, Gary R. Sisson, and Richard A. Swanson. "Cost Effectiveness: A Model for Assessing the Training Investment." Training and Development Journal 32 (January 1978): 24-27, 29.

Cummings, Oliver W. "Evaluation of Course #607184." Paper presented at the annual meeting of the Midwest Educational Research Association, Chicago, September 1984.

DeMeuse, K.P., and S.J. Liebowitz. "An Empirical Analysis of Team-Building Research." Group and Organizational Studies 6 (September 1981): 357-78.

Deming, Basil S. "A System for Evaluating Training Programs." Personnel 56 (November-December 1979): 33-41.

Dopyera, John, and Louise Pitone. "Decision Points in Planning the Evaluation of Training." Training and Development Journal 36 (May 1983): 66-71.

Elsbree, Asia Rial, and Christine Howe. "An Evaluation of Training in Three Acts." Training and Development Journal 31 (July 1977): 10-14; (August 1977): 12-19; (September 1977): 30-36.

Fauley, F.E. "Cost Models: A Study in Persuastion." Training and Development Journal 29 (June 1975): 3-8.

Fiero, Janet D. "Success Breeding Success: Motorola's Implementation of JOQI as a Quality Resource." Paper presented at Impro 84 Quality Improvement Conference, 1984.

Foshay, Wellesley R. "QA and QC: A Training Vendor's View of the Formative/Summative Evaluation Distinction." Performance and Instruction Journal 23 (December 1984): 15-17.

Gaskell, R., and R. Svenson. "Exchanging Ideas on Evaluation: Training Operational Reviews -- A Multipurpose Tool." NSPI Journal 17 (March 1978): 20-21.

Geroy, Gary D., and Richard A. Swanson. "Forecasting Training Costs and Benefits in Industry." Journal of Epsilon Pi Tau. In press.

Gilbert, Thomas F. Human Competence. Engineering Worthy Performance. New York: McGraw-Hill, 1978.

Hamblin, A.C. Evaluation and Control of Training. London: McGraw-Hill 1974.

Hand, Herbert H., and John W. Slocum, Jr. "A Longitudinal Study of the Effects of a Human Relations Training Program on Managerial Effectiveness." Journal of Applied Psychology 56 (1972): 412-17.

Hunter, Robert H., and Oliver Cummings. "Specific Risk Analysis." Paper presented at the annual meeting of the National Society for Performance and Instruction, Dearborn, Michigan, April 1983.

Hunter, Robert H., and Robert I. Nassauer. "Implementing Training Course Development Evaluation in a Corporate Setting: A Case Study." Paper presented at the annual meeting of the National Society for Performance and Instruction, 1982.

Jones, A., and J. Moxham. "Costing the Benefits of Training." Personnel Management 4 (1969): 22-28.

Kaufman, Roger. "A Holistic Planning Model." Performance and Instruction Journal 22 (October 1983): 3-12.

Kearsley, Greg, and Terry Compton. "Assessing Costs, Benefits and Productivity in Training Systems." Training and Development Journal 31 (January 1981): 52-61.

Kettenring, J.R., W.H. Rogers, M.E. Smith, and J.C. Warner. "Cluster Analysis Applied to the Validation of Course Objectives." Journal of Educational Statistics 1 (Spring 1976): 39-57.

Kirkpatrick, D.L. "Evaluating In-House Training Programs." Training and Development Journal 32 (September 1978): 6-9.

Kirkpatrick, D.L. "Evaluation of Training." In Training and Development Handbook, edited by R.L. Craig. New York: McGraw Hill, 1976.

Levin, Henry M. "Cost Analysis." Chapter 1 in New Techniques for Evaluation, edited by Nick Smith, 13-70. Berverly Hills, CA: Sage Publications, 1981.

Levin, Henry M. "Cost Effectiveness Analysis in Evaluation Research." Chapter 5 in Handbook of Evaluation Research Vol. 2, edited by Marcia Guttenberg and Elmer L. Strueving, 89-122. Beverly Hills: Sage Publications, 1975.

Lien, L. "Reviewing Your Training and Development Activities." Personnel Journal 58 (1979): 791-94.

Lyons, J. Daniel. "Measuring Effectiveness: Quality Control of Training." Human Resources Research Organization (HumRRO), Professional Paper (July 1972): 16-72.

MacFarland, R., and K. Keeler. "Assessing the Value of a Training Course." Paper presented at the annual meeting of the National Society for Performance and Instruction, Washington, DC, April 1975.

Markle, Susan M. "Evaluating Instructional Programs: How Much is Enough?" NSPI Journal 18 (February 1979): 22-24.

Martin, Vicky G. "Buying a Training Workshop: How to Tell the Terrible from the Terrific." Training (July 1983): 32-35.

Mikesell, John L., John A. Wilson and Wendell Lawther. "Training Program and Evaluation Model." Public Personnel Management (November-December 1975): 405-411.

Mirabal, Thomas E. "Forecasting Future Training Costs." Training and Development Journal 32 (July 1978): 78-87.

Moore, Elizabeth Randall. "Competency-Based Training Evaluation." Training and Development Journal (November 1984): 92-94.

Morrison, Gary R. "Developing Evaluation Programs for Industry." Paper presented at the annual meeting of the American Education Research Association, SIG/Training in Business and Industry Symposium, Los Angeles, California, April 1981.

Nickols, Frederick W. "Finding the Bottom Line Payoff for Training." Training and Development Journal (December 1979): 54-63.

Odiorne, G.S. "Training to be Ready for the '90's." Training and Development Journal 34 (1980): 12-20.

Odiorne, George S. "The Need for an Economic Approach to Training." Training and Development Journal (June 1979): 32-40.

Olivas, L. "Auditing Your Training and Development Function." Training and Development Journal 34 (1980): 60-64.

Parker, Barbara L. "Summative Evaluation in Training and Development: A Review and Critique of the Literature, 1980 through 1983." Project Number Four, Training and Development Research Center, University of Minnesota, October 1984. Mimeograph.

Phillips, Jack. "Evaluating Outside Resources." Chapter 11 in Training Evaluation and Measurement Methods. Houston: Gulf Publishing, 1983.

Putnam, Anthony O. "Pragmatic Evaluation." Training and Development Journal 34 (October 1980): 36-40.

Reber, Robert A., and Jerry A. Wallin. "The Effects of Training, Goal Setting, and Knowledge of Results on Safe Behavior: A Component Analysis." Academy of Management Journal 27 (1984): 544-60.

Rogers, Mark A., and Sam Volpe. "Evaluation of Vendor Courses." Performance and Instruction Journal 23 (1984): 18-22.

Rosentreter, Gary. "Evaluation Training by Four Economic Indices." Adult Education 29 (1979): 234-41.

Ross, Paul C. "Exchanging Ideas on Evaluation: 4. A Model of Evaluating Off-the-Shelf Training" NSPI Journal 16 (1977): 4-5.

Rossi, Peter H., Howard E. Freeman, and Sonia R. Wright. "Measuring Efficiency." Chapt. 8 in Evaluation: A Systematic Approach. Beverly Hills: Sage Publications, 1979.

Rothwell, William J. "How to Conduct a Real Performance Audit." Training (June 1984[a]): 45-49.

Rothwell, William J. "Strategic Needs Assessment." Performance and Instruction Journal 23 (July 1984[b]): 19-20.

Scherman, I.A. "Training: Directions for the '80's." Training and Development Journal 34 (January 1980): 50-55.

Schumacher, S.P., R.A. Kulp, and C.E. Childs, Jr. Bell System, 1975. Mimeograph.

Smith, Martin E. "Trends in Training Evaluation." Paper presented at the annual meeting of the National Society for Performance and Instruction, Atlanta, April 1984.

Smith, Martin E. "Field Research Designs." Performance and Instruction 21 (February 1982[a]): 27-29.

Smith, Martin E. "An Industrial Application of Meta-Evaluation." Evaluation and Program Planning 5 (1982[b]): 99-110.

Smith, Martin E. "A Training Information System. NSPI Journal 21 (July-August 1982[c]): 20-25.

Smith, Martin E. "Evaluating Training Operations and Programs." Training and Development Journal (October 1980): 70-78.

Smith, Martin E. "Training Cost Accounting at New England Telephone." New England Telephone Company, 1980. Mimeograph.

Smith, Martin E. "Exchanging Ideas on Evaluation: 15. New England Telephone's Training Management Operational Review Plan." NSPI Journal 18 (July 1979): 44-47.

Smith, Martin E. "An Illustration of Evaluating Post-Training Job Performance." Improving Human Performance Quarterly 8 (1979): 181-201.

Smith, Martin E. "Evaluation of the Attendance Administration Course." New England Telephone Company, October 1975. Unpublished report.

Smith, Martin E., and Al Corbett. "Exchanging Ideas on Evaluation: Pricing and Benefits of Training." NSPI Journal 16 (1977): 16-18.

Smith, Martin E., and R. Marcinuk. "Training Cost Accounting at New England Telphone." In The Nature and Extent of Employee Training and Development, edited by R.L. Craig. Washington, D.C.: American Society for Training and Development, 1982.

Smith, Martin E., M. O'Callaghan, A.J. Corbett, B. Morley, and I.L. Kamradt. "An Example of Meta-Evaluation from Industry." Improving Human Performance Quarterly 5 (1976): 168-182.

Snyder, David. "Procedures and Criteria for Assessing Training Programs." Paper presented at the annual meeting of the National Society for Performance and Instruction, Atlanta, Georgia, April 1984.

Spencer, Lyle M., Jr. "How to Calculate the Costs and Benefits of an HRD Program." Training 21 (July 1984): 40-51.

Stufflebeam, Daniel L., and William J. Webster. "An Analysis of Alternative Approaches to Evaluation." Educational Evaluation and Policy Analysis 2 (May-June 1980): 5-20.

Swanson, Richard A., and Gary D. Geroy. "Forecasting the Economic Benefits of Training." Training and Development Research Center, University of Minnesota, April 1984. Mimeograph.

Temkin, Sanford. "Making Sense of Benefit-Cost Analysis and Cost-Effectivenss Analysis." Improving Human Performance: A Research Quarterly 3 (Summer 1974): 39-48.

Tennyson, Robert D. "The Role of Evaluation in Instructional Development." Educational Technology 16 (September 1976): 17-24.

Thomas, W. "How to be Successful as a Consultant to Industry." Instructional Innovation 26 (1981): 22-23.

Thompson, John T. "How to Develop a More Systematic Evaluation Strategy." Training and Development Journal 32 (July 1978): 88-93.

Tracey, W.R. Evaluating Training and Development Systems. New York: American Management Association, 1968.

Weinstein, Laurence M. "Colleting Training Cost Data." Training and Development Journal 36 (August 1982): 31-34.

Westgaard, Odin. "Auditing Operations: Performance Technology and the Bottom Line." Performance and Instruction Journal 23 (December 1984): 10-14.

Zigon, Jack. "Expanding Your Impact on Employee Performance." Paper presented at the annual meeting of the National Society for Performance and Instruction, Atlanta, Georgia, April 1984.

APPENDIX

METHODS FOR TRAINING ECONOMIC IMPACT ANALYSIS

MODEL/LABEL: Return on Investment

PROPOSER/PROPONENT: Barta (1982)

PURPOSE: Justify training investment.

Most accurate because of attention to time value of money.

USE: Examine one project in isolation to judge potential.

BASIC INPUTS/ASSUMPTIONS: Life (time) of training projects.

Amount of investment.

Cash flow after expenses.

FORMULA/METHOD: Find interest rate where present worth of all cash flow equals zero.

OUTCOME GUIDELINES: Compare ROI percent to company requirement, i.e., high or low is relative.

APPENDIX

METHODS FOR TRAINING ECONOMIC IMPACT ANALYSIS

MODEL/LABEL: Payback Period

PROPOSER/PROPONENT: Barta (1982)

PURPOSE: Length of time needed to break even.

USE: Initial examination at a questionable investment.

BASIC INPUTS/ASSUMPTIONS: Investment expenses.

Annual Savings.

FORMULA/METHOD: Payback period (in years) =

$$\frac{\text{Investment}}{\text{Annual Savings}}$$

OUTCOME GUIDELINES: If result is less than 1, then good potential.

APPENDIX

METHODS FOR TRAINING ECONOMIC IMPACT ANALYSIS

MODEL/LABEL: Human Capital

PROPOSER/PROPONENT: Odiorne (1979)

PURPOSE: Derive tangible and intangible benefits of training.

Derive standards for rates of return.

USE: Decide on where to allocate money.

Decide whether or not training is worth it.

BASIC INPUTS/ASSUMPTIONS: Classify costs into Direct, Indirect and Opportunity.

Classify costs above into:
1. Profit Improvement
2. Human Working Capital
3. Investment in Human Capital

FORMULA/METHOD: Classify economic objectives for each part of organization.
1. (above) immediate training
2. (above) training having 1 to 3 year payback, e.g., upgrade range of skills
3. (above) up to 10 years to return, e.g., management development

OUTCOME GUIDELINES: Answer four questions:
1. How much money is needed for long-range expenditure?
2. What is the supply of funds for human capital?
3. How is training expense to be rationed among alternatives?
4. What timing is needed for funds request?

APPENDIX

METHODS FOR TRAINING ECONOMIC IMPACT ANALYSIS

MODEL/LABEL: Measurement-Based Analysis

PROPOSER/PROPONENT: Nickols (1979)

PURPOSE: Methods for deciding where to intervene (foresight analysis).

Opposite of evaluation.

Build a model.

Connect organizational means (activities) to ends.

USE: Locate where intervention is needed and what payoff it is likely to have.
Select appropriate techniques/methods for intervention.

BASIC INPUTS/ASSUMPTIONS: Assume: a quantitative system of measuring basic components.
Trace inputs thru processes until they are transformed to outputs.

FORMULA/METHOD: Construct a map for organization means and ends.

Repeatedly ask 3-question sequence:
1. What is the measure?
2. How is it calculated?
3. What are its input variables?

Take result of #3 and repeat.

Stop when input variables are direct product of person's activity.

OUTCOME GUIDELINES: Identify standards for variables at each level and compare to actual values.

Quantifies cost of problem (in relation to organization standards).

Quantifies value of solution.

Separates environmental effects from individual behavior.

APPENDIX

METHODS FOR TRAINING ECONOMIC IMPACT ANALYSIS

MODEL/LABEL: Life Cycle

PROPOSER/PROPONENT: Kearsley & Compton (1981)

PURPOSE: Examine all phases of costs--start-up, operational, steady state and transition to new program.

USE: Compare costs of 2 or more training approaches at a given point in time.

A method to look at savings.

BASIC INPUTS/ASSUMPTIONS: Given approach x & y
Total costs savings = Total Costs x - Total Costs y

Transition costs will be dual

FORMULA/METHOD: Total Life Cycle Costs (t) =

Total Start-Up Costs (s) + Total Transition Costs (n) + Total Steady State Costs (m)

OUTCOME GUIDELINES: The shorter the transition period and the longer the steady state period, the greater the savings to be realized.

APPENDIX

METHODS FOR TRAINING ECONOMIC IMPACT ANALYSIS

MODEL/LABEL: Productivity Function Analysis

PROPOSER/PROPONENT: Kearsley and Compton (1981)

PURPOSE: Weigh both efficiency (costs) and effectiveness to look at proficiency produced for a given set of training resources.

USE: Can predict amount of money spent on a given training activity.

Determine maximum productivity to be gained by a given training situation.

Obtain productivity curves to compare across alternatives.

BASIC INPUTS/ASSUMPTIONS: Comes from classic economic theory:
$O = aL^bC^c$
O = production output
L = amount of labor
C = amount of capital
a = constant
b & c = coefficients

There is always some point where a given training approach reaches its limits.

Useful only for large training programs with significant budgets.

FORMULA/METHOD: Amount of training accomplished (1) = number of trainers x skills of trainers x training procedures x training technology.
(1) = number of graduates

From student perspective:
Amount of Learning Accomplished = nature of learner profile x nature of instructional presentation x amount of practice.

OUTCOME GUIDELINES: Increase in any term of function increase output but at diminishing rate.

If any term is zero, the output is zero.

Increasing all terms by the same factor increases the output by the same factor.

APPENDIX

METHODS FOR TRAINING ECONOMIC IMPACT ANALYSIS

MODEL/LABEL: Economic Benefits Forecasting (Performance Value Analysis)

PROPOSER/PROPONENT: Geroy & Swanson (1984)

PURPOSE: A type of cost-benefit analysis

A comparative analysis of training costs and benefits (performance value)

USE: Compare two or more methods of training for cost/benefit in a forecasting mode.

BASIC INPUTS/ASSUMPTIONS: In instructional training, performance is 50% of goal, i.e., 50% opportunity costs.

A base time period must be chosen equal to the longest time to bring trainee to performance goal.

Figures are projected with assistance from management.

Worksheets available for both costs and performance value.

FORMULA/METHOD: List all training costs--staff, consultants, materials, external support costs, trainee, facilities, tuition, fees.

Specify total performance or value of performance unit.

Subtract cost from performance value = net benefits.

OUTCOME GUIDELINES: Assist in making strategic plans.

An example application is given.

APPENDIX

METHODS FOR TRAINING ECONOMIC IMPACT ANALYSIS

MODEL/LABEL: Cost-Benefit Analysis

PROPOSER/PROPONENT: Levin (1981); Kearsley & Compton (1981); Temkin (1974); Rossi, Freeman & Wright (1979)

PURPOSE: Select one alternative from a number of alternatives with both costs and benefits expressed in monetary terms.

USE: Can compare across projects and target groups, similar objectives not required.

Can be used to perform sensitivity analysis (determine importance by changing parameters).

Make subjective judgments on training effectiveness explicit.

BASIC INPUTS/ASSUMPTIONS: Requires costs and benefits expressed in quantitative (monetary) terms.

Convert tangible and intangible benefits to monetary terms.

Requires separation of training system parameters from training benefits and operational benefits.

FORMULA/METHOD: For tree structure of benefit linkage:

1. assign values for strength of causal links
2. $B^j = V^i \times B^i$ (B^i = parameter value)
3. B^j = j-th benefit; V^i is value of Bi, k = #benefits affecting B

In general, form ratio generally expressed as B/C and compare.

In present value terms: $V^1 = (B^{it} - C^{it})/(1 + i)^t$ t is time in years project is analyzed, B^{it} is benefit of alternative 1 in year t, i is discount factor, discount factor is like reverse interest rate.

OUTCOME GUIDELINES: Assumptions require stretching reality in many cases. Most advocate use only as screening medium or forecasting tool.

APPENDIX

METHODS FOR TRAINING ECONOMIC IMPACT ANALYSIS

MODEL/LABEL: Cost-Effectiveness Analysis

PROPOSER/PROPONENT: Temkin (1974); Levin (1981); Rossi, Freeman and Wright (1979)

PURPOSE: Select one alternative from a number of alternatives, each designed to meet one or more objectives.

USE: For given effectiveness level, choose alternatives to minimize costs.

For given cost, choose alternative that maximizes effectiveness.

BASIC INPUTS/ASSUMPTIONS:

Assumptions: Decision alternatives (more than one)

Cost analysis accompanied by evaluation of effectives of each.

Programs must have identical or similar goals.

Inputs Required: Set of objectives.

Priority assignment to objectives

Structure that relates training to objectives.

Indicies of relative importance to alternatives.

Performance indicators for evidence of actual contribution of alternatives to objectives of system.

FORMULA/METHOD: Costs must be made comparable--use present value (PV) formula

PV

X^T = cost in year t; i = interest rate (usually 5 to 10%)

n = final year

Effectiveness for activity i = $p^i V^i$; p^i = performance index V^i = value of activity.

OUTCOME GUIDELINES: A structure to choose among alternatives

Common measures of effectiveness are used or interpretation is impossible.

Cannot establish whether or not investment is worthwhile.

RECENT TITLES
IN THE ERIC/TME REPORT SERIES

#90 - Assessing Higher Order Thinking Skills, by C. Phillip Kearney, and Others. 4/86, $7.50.

#89 - Microcomputer Programs for Educational Statistics: A Review of Popular Programs, by Paul M. Stemmer and Carl F. Berger. 3/86, $7.50.

#88 - Computer-Assisted Test Construction. A State of the Art, by Tse-chi1 Hsu and Shula F. Sadock. 12/85, $7.50.

#87 - The Statewide Assessment of Writing, by Peter Afflerbach. 8/85, $7.50.

#86 - The Effects of Testing on Teaching and Curriculum in a Large Urban School District, by Floraline Stevens. 12/84, $6.00.

#85 - Reporting Test Scores to Different Audiences, by Joy A. Frechtling and N. James Myerberg. 12/83, $7.00.

#84 - Assessment of Learning Disabilities, by Lorrie A. Shepard. 12/82, $6.50.

#83 - Statistical Methodology in Meta-Analysis, by Larry V. Hedges. 12/82, $7.00.

#82 - Microcomputers in Educational Research, by Craig W. Johnson. 12/82, $8.50.

#81 - A Bibliography to Accompany the Joint Committee's Standards on Educational Evaluation, compiled by Barbara M. Wildemuth. 12/81, $8.50.

#80 - The Evaluation of College Remedial Programs, by Jeffrey K. Smith and others. 12/81, $8.50.

#79 - An Introduction to Rasch's Measurement Model, by Jan-Eric Gustafsson. 12/81, $5.50.

#78 - How Attitudes Are Measured: A Review of Investigations of Professional, Peer, and Parent Attitudes toward the Handicapped, by Marcia D. Horne. 12/80, $5.50.

#77 - The Reviewing Processes in Social Science Publications: A Review of Research, by Susan E. Hensley and Carnot E. Nelson. 12/80, $4.00.

#76 - Intelligence Testing, Education, and Chicanos: An Essay in Social Inequality, by Adalberto Aguirre Jr. 12/80, $5.50.

#74 - Intelligence, Intelligence Testing and School Practices, by Richard DeLisi. 12/80, $4.50.

#72 - Methods of Identifying Gifted Minority Students, by Ernest M. Bernal. 12/80, $4.50.